The Gamma-Ray Burst Afterglow Modeling Project: Foundational Statistics and Absorption & Extinction Models

Adam Somers Trotter

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics & Astronomy.

Chapel Hill
2011

Approved by:

Daniel E. Reichart, Advisor

Gerald Cecil, Committee Member

Arthur Champagne, Committee Member

Christopher Clemens, Committee Member

Fabian Heitsch, Committee Member

Abstract

ADAM SOMERS TROTTER: The Gamma-Ray Burst Afterglow Modeling Project: Foundational Statistics and Absorption & Extinction Models. (Under the direction of Daniel E. Reichart.)

The Gamma-Ray Burst (GRB) Afterglow Modeling Project (AMP) will model, in a statistically sound and self-consistent way, every GRB afterglow observed since the first detection in 1997, using all available radio, infrared, optical, ultraviolet and X-ray data. The result will be a catalog of fitted empirical model parameters describing the intrinsic afterglow emission, and extinction due to dust and absorption due to gas along the line of sight to the GRB. This ever-growing catalog of fitted model parameters will allow us to infer the astrophysical properties of GRBs and their environments, and to explore their variety and evolution over the history of the universe. First, I present a new, broadly applicable statistical technique, the TRF statistic, for fitting model distributions to data in two dimensions, where the data have intrinsic uncertainties in both dimensions, and extrinsic scatter in both dimensions that is greater than can be accounted for by the intrinsic uncertainties alone. I demonstrate the properties of the TRF statistic, which is invertible but not scalable, and present an algorithm for obtaining an optimum scale for fits to a given data set. I then apply the TRF statistic to observations of interstellar extinction of stars along various Milky Way and Magellanic Cloud lines of sight, and to observations of Lyα forest flux deficits in quasars, to construct a comprehensive empirical model for extinction due to interstellar dust in the source frame and in the Milky Way, and absorption due to gas in the source frame and in the intergalactic medium. Combined with theoretical models of synchrotron emis-

sion from GRB jets, the resulting parameterization provides a framework for modeling the observed emission from most GRB afterglows. Furthermore, the extinction and absorption models are broadly applicable, in that they may be used to model observations of any extragalactic point source of radiation. Finally, I describe the results of model fitting to NIR, optical and X-ray observations of the afterglow of GRB 090313, which exercises all aspects of the AMP modeling framework presented in this thesis.

In Honor of My Parents,
The Rev. J. Albert Trotter & Mrs. Gaynelle C. Trotter,

And of My Grandmothers,
Mrs. Ethel R. Trotter & Mrs. Eleanor F. Chappell,

And in Loving Memory of My Grandfathers,
Mr. Franklin H. Trotter & Mr. Robert M. Chappell,

And of My Great-Grandfather, in Whose Footsteps I Follow,
Mr. Robert G. Fitzgerald, UNC–Chapel Hill, Class of 1915

Para
Vicente Rosario Meléndez

Acknowledgments

I gratefully acknowledge the financial, scientific and moral support of: The NASA/North Carolina Space Grant Graduate Research Fellowship Program; The U.S. Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship Program; The Sigma Xi Grant-in-Aid of Research Program; The Department of Physics & Astronomy at UNC–Chapel Hill; Prof. Gerald Cecil (UNC-Chapel Hill); Prof. Don Smith (Guilford College); and especially past and present members, affiliates and friends of the Reichart Lab/Skynet Robotic Telescope Network team: Brad Barlow, Matt Bayliss, Summers Brennan, Rebecca Egger, Andrew Foster, Joshua Haislip, Kevin Ivarsen, Aaron LaCluyzé, Justin Moore, Melissa Nysewander, Apurva Oza, Rachel Rosen, Mark Schubel, Eric Speckhard, Jana Stýblová & Elise Weaver. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester.

Few graduate students in my experience have been so fortunate as to have an advisor as attentive, empowering and enthusiastic as Dan Reichart. I am extremely grateful for his constant support and guidance during the research for and writing of this thesis, and in nearly every aspect of my student and professional life. It has been a great honor to work together with him and his remarkable team, and I look forward to continuing to do so now that this thesis is complete. Thank you, Dan.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

"The thing can be done," said the Butcher, "I think.
The thing must be done, I am sure.
The thing shall be done! Bring me paper and ink,
The best there is time to procure."
— Lewis Carroll, *The Hunting of the Snark*, Fit the Fifth

1.1 AMP: The GRB Afterglow Modeling Project

The Gamma-Ray Burst (GRB) Afterglow Modeling Project (AMP) will model, in a statistically sound and self-consistent way, every GRB afterglow observed since the first detection in 1997, using all available radio, infrared, optical, ultraviolet and X-ray data. The result will be a catalog of fitted empirical model parameters describing the intrinsic afterglow emission, and extinction due to dust and absorption due to gas along the line of sight to the GRB. This ever-growing catalog of fitted model parameters will allow us to infer the astrophysical properties of GRBs and their environments, and to explore their variety and evolution over the history of the universe. By carefully fitting models to the observed afterglow flux as a function of frequency and time, taking into account all line-of-sight sources of extinction due to dust and absorption due to gas, it is possible to recover, at the very least, an empirical description of the intrinsic emission.

If the data are of sufficient quality and coverage, it is possible to obtain, from the fitted empirical model, values for various intrinsic physical parameters of the GRB and of the circumburst medium. The fitted source-frame extinction and absorption parameters can be used to explore the nature of interstellar dust and gas, both in the vicinity of, and under the duress of, the GRB, and away from the GRB in its host galaxy.

1.2 Statistic

Before we can construct a generalized model for extinction by dust and absorption by gas for a GRB, or any point source of radiation, it is necessary to establish a solid statistical foundation. In Chapter 2, we develop a new, very generally applicable statistic (Trotter, Reichart & Foster, in preparation, hereafter TRF) for fitting model distributions to data in two dimensions, where the data have both intrinsic uncertainties (i.e., error bars) and extrinsic scatter (i.e., sample variance) that is greater than can be attributed to the intrinsic uncertainties alone. A model distribution is described by the convolution of a (possibly asymmetric) 2D probability distribution that characterizes the extrinsic scatter, or "slop", in the data in both the x- and y-dimensions with a model curve $y_c(x;\vartheta_m)$, where $\{\vartheta_m\}$ is a set of M model parameters that describe the shape of the curve. The task in fitting a model distribution to a data set is to find the values of the parameters $\{\vartheta_m\}$, and of the parameters that characterize the slop, that maximize the joint probability, or likelihood function, of the model distribution and the data set. First, we derive a very general probabilistic formulation of the joint probability for the case where both the intrinsic uncertainties and the slop are normally distributed (§2.2). We then demonstrate that there is a choice of rotated coordinate systems over which the joint probability integrals may be evaluated, and that different choices of this rotated coordinate system yield fundamentally different statistics with fundamentally different properties, and yield fundamentally different predictions for

any given model parameterization and data set (§2.3).

We show that one choice of rotated coordinates yields the traditional statistic advocated by D'Agostini (2005, hereafter D05), and another the statistic advocated by Reichart (2001, hereafter R01). The D05 statistic is *non-invertible*, i.e., it yields different model fits depending on whether one chooses to fit a model distribution to y vs. x or to x vs. y (§2.4.1). The R01 statistic *is* invertible, but as we will demonstrate, suffers from a fatal flaw in that it does not reduce to χ^2 in the 1D limiting case of zero extrinsic scatter, and zero intrinsic uncertainty in either the x- or y-dimensions (§2.4.2). We introduce a new statistic, TRF, that is both invertible and reduces to χ^2 in these 1D limiting cases. We demonstrate that the best-fit model distribution predicted by the TRF statistic, in the case of normally distributed intrinsic uncertainties and extrinsic scatter, is geometrically equivalent to the distribution that minimizes the sum of the squares of the radial distances of each data point centroid from the model curve, measured in units of the 1σ radius of the 2D convolved intrinsic and extrinsic Gaussian error ellipse (§2.3).

However, unlike the D05 statistic, the TRF statistic is not *scalable*, i.e., it yields different best-fit model distributions depending on the choice of basis for each coordinate axis (§2.4.4). We demonstrate that, in the limit of data that are entirely dominated by extrinsic scatter (i.e., zero intrinsic error bars), the predictions of TRF at extremes of scaling the x- and y-axes is equivalent to the predictions of D05 under an inversion of the x- and y-axes (§2.5). However, D05 is limited to this binary choice of inversion or non-inversion, while TRF is free to explore a continuous range of scales between the two limiting cases. Moreover, when the data are dominated by intrinsic rather than extrinsic uncertainty, the range of scales that result in physically meaningful model distributions is significantly smaller, while the predictions of D05 are the same under inversion whether the data are dominated by intrinsic or extrinsic uncertainty. In the

limit of data that are entirely dominated by intrinsic uncertainty (i.e., zero slop), this range of scales reduces to a single physically meaningful scale. We discuss these behaviors in the context of linear model distributions fit to various randomized data sets. We define a new correlation coefficient, $R^2_{\rm TRF}$ (§2.5.1), that is analogous to the well-known Pearson correlation coefficient R^2_{xy}, and describe an algorithm for arriving at an "optimum scale" for a fit with TRF to a given data set, which yields linear fits that fall approximately midway between the inverted and non-inverted D05 fits (§2.5.2). Lastly, we generalize this algorithm to non-linear fits (§2.5.3), and to fits to data with asymmetric intrinsic or extrinsic uncertainties (§2.7.2 and Appendices A & B).

1.3 Model

In Chapter 3, we describe our model for the intrinsic emission, and all sources of line-of-sight extinction due to dust and absorption due to gas, of a GRB afterglow. First, we describe the intrinsic emission model (§3.2), based on the standard blast wave model (e.g., Mészáros & Rees, 1997; Sari et al., 1998; Piran, 1999; Chevalier & Li, 1999; Granot et al., 2000; Mészáros, 2002, and references therein). We assume that the intrinsic emission of an afterglow is due to synchrotron radiation from shocks formed when a relativistic jet emerging from the GRB progenitor collides with, and is decelerated by, the circumburst medium. This blast wave model predicts intrinsic emission that is described by smoothly broken power laws in frequency and time. By fitting the spectral and temporal power-law indices of the intrinsic afterglow emission (accounting properly for extinction and absorption), it is possible to infer: the power-law index p of the energy distribution of the shocked electrons; and the power-law index k of the radial density profile of the circumburst medium, which is typically modeled as $n(r) \propto r^k$, where $k = 0$ corresponds to a constant-density medium, and $k = -2$ to an idealized stellar wind-blown medium. If the data are of sufficient quality and coverage,

it is possible to obtain, from the fitted empirical model, values for additional intrinsic physical parameters of the GRB and its environment, including: the total isotropic-equivalent energy, $E_{\rm iso}$, of the burst; the fraction of this energy that is contained in electrons, ϵ_e, and in the magnetic field, ϵ_B; and the circumburst density n (in the case of a constant-density medium) or A_* (in the case of a wind-blown or other non-constant medium).

We next apply the TRF statistic to fit empirical model distributions to the distribution of and correlations between observed interstellar extinction parameters in the Milky Way and Magellanic Clouds (§3.3), and to observed IGM Lyα forest flux deficits as a function of redshift in a sample of QSOs (§3.4.2). The fitted probability distributions of the secondary model parameters describing these distributions define a set of priors that, in turn, constrain the values of the parameters comprising a comprehensive model for sources of line-of-sight extinction due to dust and absorption due to gas of a GRB afterglow, or of any extragalactic point source of radiation. These model fits, and the GRB model fits themselves, are all obtained using the highly flexible genetic algorithm software package, *Galapagos* (§2.6.3; Foster & Reichart 2011, in preparation). Only when all significant sources of extinction and absorption are taken into account is it possible to accurately model the intrinsic emission of a GRB afterglow. Furthermore, the fitted values of extinction and absorption parameters are physically interesting in their own right, providing information about the amount and properties of source-frame dust and gas. By carefully modeling changes in extinction as a function of time, it may be possible to detect signatures of dust modification by the burst (at early times), and illumination of different populations of dust as the GRB jet expands laterally (at later times).

We discuss, in a very general sense, a number of mistakes that are commonly made when fitting models to data, and describe ways to avoid these pitfalls (§3.5.1). We

then describe our method for fitting the combined emission, extinction and absorption models to photometric observations of GRB afterglows (§3.5.2). Finally, in Chapter 4, we describe fits to NIR, optical and X-ray observations of GRB 090313, which exercises nearly all aspects of the AMP emission, extinction and absorption models.

Chapter 2

Statistic

"The method employed I would gladly explain,
While I have it so clear in my head
If I had but the time and you had but the brain—
But much yet remains to be said."
— Lewis Carroll, *The Hunting of the Snark*, Fit the Fifth

2.1 Overview of Bayesian Statistics

As we will describe fully in Chapter 3, a full description of a gamma-ray burst's emission, extinction and absorption as functions of time and frequency requires a large number of model parameters. If all these parameters were free to take on any numerical value, fitting a model to a given set of observations would be an exercise in futility; the model would have too many degrees of freedom, multiple degeneracies, and any given set of best-fit parameter values would have little physical or statistical significance. But the volume of this parameter space can be reduced if we take advantage of prior information, obtained by analysis of other sets of observations of the phenomena in question. The use of prior information to inform and constrain the predictions of a model fit to data is the heart of Bayesian analysis.

We begin with a brief overview of Bayesian analysis; see Reichart (2001) for a more

complete discussion. Let D represent a set of observed data, H a set of parameters defining a hypothetical model, and I any prior information that is available. Our task is to determine the probability[1] of a particular model H, conditional on the observed data D and prior information I, i.e., $p(H|DI)$.

Bayes' theorem states that:

$$p(H|DI) = \frac{p(D|HI)p(H|I)}{p(D|I)}\,. \tag{2.1}$$

The quantity $p(H|DI)$ is called the *posterior* probability function, $p(D|HI)$ is known as the *likelihood* function, and $p(H|I)$ is called the *prior* probability function. The quantity in the denominator, $p(D|I)$, is determined by the requirement that the posterior probability function be normalized. In these terms, Equation 2.1 states that *the posterior is proportional to the likelihood times the prior.*

The likelihood function $p(D|HI)$, which is usually notated $\mathcal{L}$, is the conditional probability of obtaining a given set of observations, given a hypothetical parameterized model H; how to compute $\mathcal{L}$ is the primary subject of this chapter. The prior function $p(H|I)$ describes how any pre-existing information constrains the values of each of the model parameters. In some cases, there is no a priori reason to prefer one value of a given parameter over another, except possibly to restrict it to a physically meaningful range; in this case, the prior will be a constant over this range (a *flat prior*). In other cases we can define, based on previous measurements, a functional form (e.g., Gaussian) for the prior.

As a hypothetical example, consider a model that includes Hubble's Constant H_0 as one of its parameters. In the 1980's, a reasonable value for Hubble's Constant, based on all the cosmological literature, might have been $H_0 = 60 \pm 20$ km s^{-1}Mpc^{-1}, while

[1]Technically, this is a *probability density* over the set of parameters that characterize H, relative to the universe of all other possible hypothetical models H' and their sets of parameters (see §2.6).

today, a better estimate is $H_0 = 72 \pm 8$ or so (e.g., Freedman et al., 2010). If one were conducting model fits back in the 1980's, and simply held Hubble's Constant fixed at $H_0 = 60$, this value would propagate through the model, and any fits obtained would be biased, relative to what would be obtained with $H_0 = 72$. Further, by holding H_0 constant, the confidence intervals obtained by fitting the other parameters to the data would be *overly narrow*; none of the results might be consistent with a value $H_0 = 72$. If, on the other hand, H_0 were allowed to be entirely free, with a flat prior between $0 \leq H_0 < \infty$, H_0 would be free to take on any value that the data and other model parameters allowed, and the confidence intervals of other fitted parameters would be *overly broad*. The proper approach, in Bayesian analysis, would be to assign a Gaussian prior to H_0; in our 1980's example, the peak of the Gaussian would be at $H_0 \approx 60$, and its 1σ width would be $\approx$20. The value of H_0 that is most consistent with the data and the rest of the model that you are fitting is still free to deviate from the peak of the prior, but with a penalty to the total fitness that increases as H_0 drifts further from that peak. In a sense, the Gaussian prior on H_0 is analogous to an additional data point. The confidence intervals of the other fitted parameters will be constrained, in part, by the peak and width of the prior on H_0. If the same fit to the same data were performed again today, the peak and width of the Gaussian H_0 prior would have to be revised to the current $H_0 = 72 \pm 8$; the new best-fit parameter values would be different, and their confidence intervals would be narrower; but they would, at least, lie within the confidence intervals of the earlier fit.

In our GRB afterglow models, some parameters have either flat or roughly Gaussian priors on their values. These are relatively straightforward to define and to implement. For example, our model for extinction due to dust in the Milky Way includes the parameter $R_V^{\rm MW} \equiv A_V/(A_B - A_V)$, where A_V and A_B are the extinction magnitudes in the V and B bands (see §3.3). The canonical value of this parameter along Milky Way

lines of sight is $R_V^{\rm MW} = 3.1$. Often, in modeling extinction, others hold $R_V^{\rm MW}$ fixed at this value. However, in reality, it has been measured to be as low as $\approx$2.5 and as high as $\approx$6. In §3.3, we describe our analysis of R_V measured along 413 Milky Way lines of sight, from which we obtain an asymmetric Gaussian prior on $\log R_V^{\rm MW} = 0.423^{+0.082}_{-0.010}$ (Figure 2.1); this is what we use so as to not unnecessarily *underestimate* ($R_V^{\rm MW} = 3.1$) or *overestimate* (flat prior on $R_V^{\rm MW}$) the confidence intervals on other fitted parameters.

But there are other, slightly more complicated cases where prior information may exist in the form of correlations between two (or more) model parameters. Consider the following simple example: In §3.3 we present an empirical dust extinction model by Fitzpatrick & Massa (1988) for use in the ultraviolet. Two of the model parameters (which we need not define here) are c_1 and c_2. Figure 2.2 is a plot of c_1 vs. c_2 measured along the lines of sight to 441 stars in the Milky Way and Magellanic Clouds. Clearly, the measured values are correlated. We can describe this correlation by a slope m_{c_1} and an intercept b_{c_1}. The extinction parameter c_1 can then be replaced in our GRB model by two new model parameters (m_{c_1}, b_{c_1}), which together with the (free) parameter c_2 fully specify the value of c_1. We analyze linear model fits to the c_1 vs. c_2 data to obtain probability distributions for m_{c_1}, b_{c_1}; these probability distributions, in turn, become priors in our GRB model fit (Figure 2.3). Hence, instead of two free parameters, we have one free parameter and two parameters that are tightly constrained by priors. Others use the best-fit values of b_{c_1} and m_{c_1} to define c_1 as a function of c_2. This approach, like the $R_V^{\rm MW} \equiv 3.1$ assumption discussed earlier, leads to underestimated confidence intervals in the other fitted model parameters.

2.2 Fitting a Model to Data in Two Dimensions

We now turn to the very general problem of fitting a model to a two-dimensional set of data points. Each data point may have *intrinsic scatter*, or *statistical uncertainty*, in

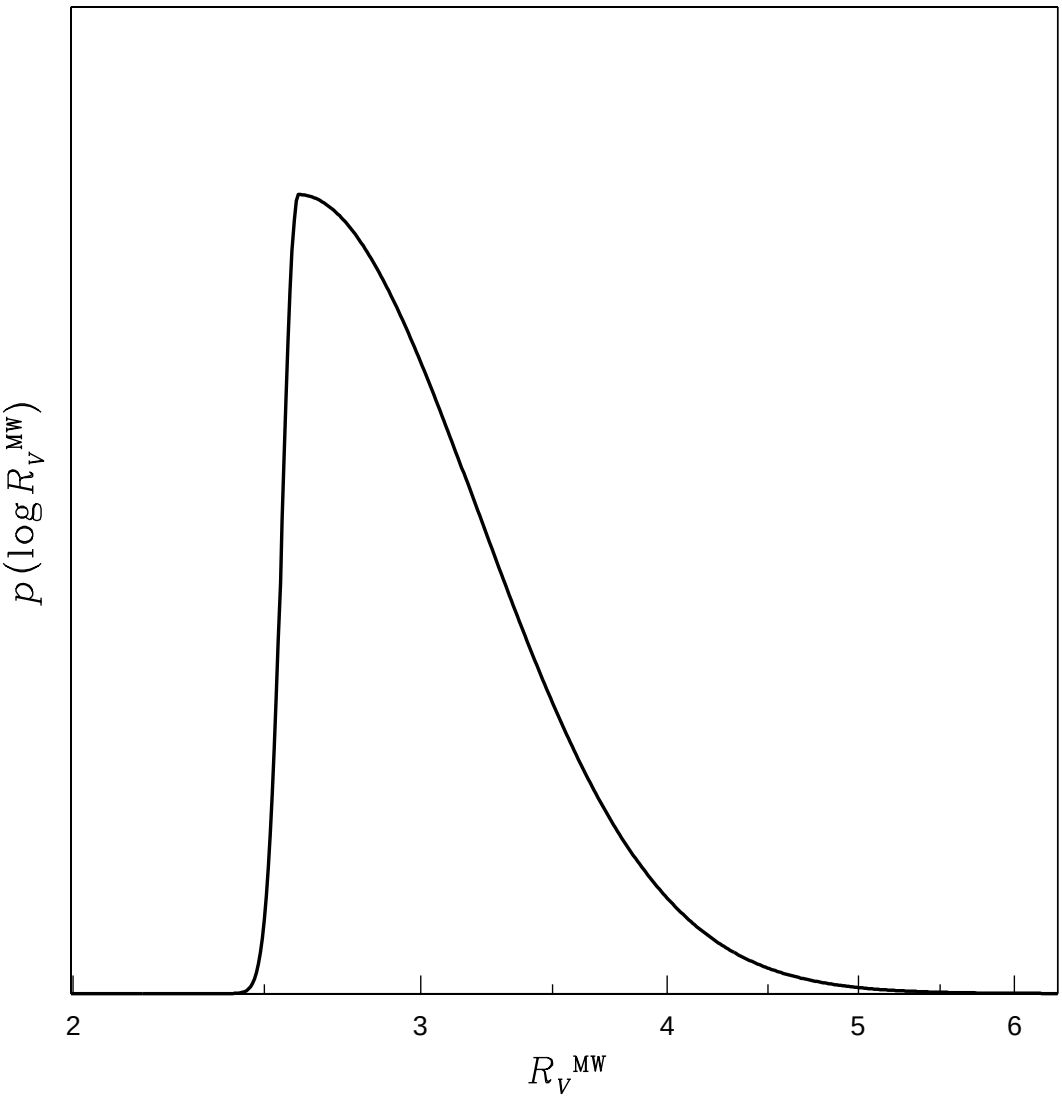

Figure 2.1 Asymmetric Gaussian prior on Milky Way extinction parameter $\log R_V^{\rm MW}$ (from §3.3).

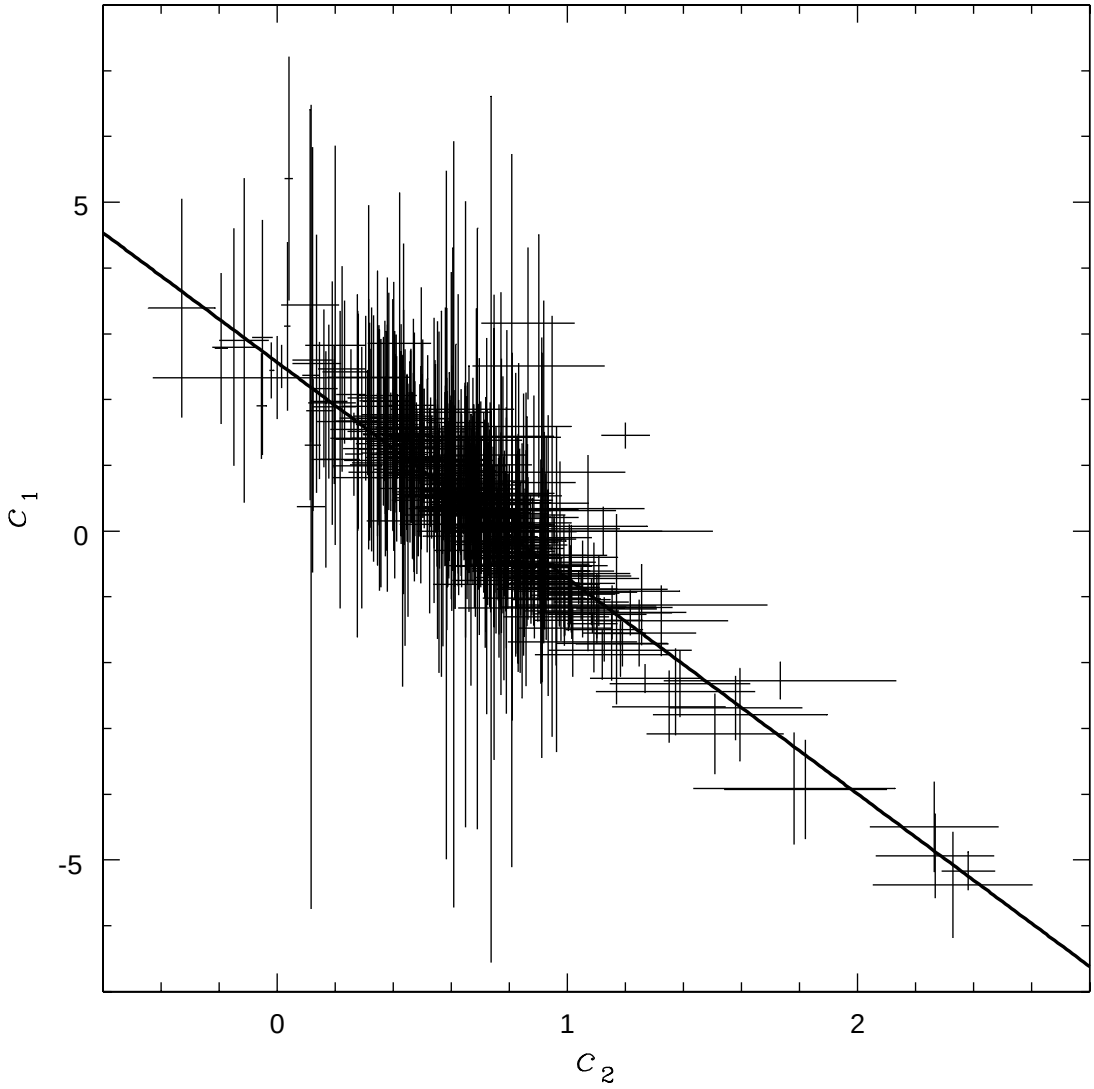

Figure 2.2 Linear correlation of extinction parameters c_1 vs. c_2 (from §3.3). For simplicity, we do not show or address the extrinsic scatter in these data here.

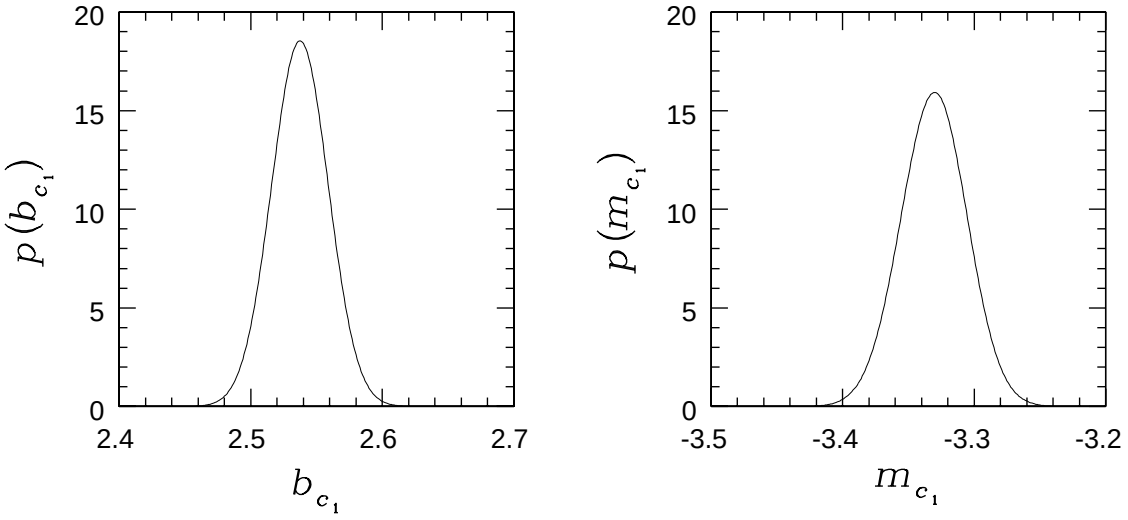

Figure 2.3 Prior probability distributions of linear correlation parameters b_{c_1} and m_{c_1}. In the GRB model, parameter c_1 is a function of these two parameters and c_2: $c_1 = m_{c_1} c_2 + b_{c_1}$ (from §3.3).

both dimensions, which can vary from point to point. A data point's intrinsic scatter is commonly referred to as its *error bars*. The data set as a whole may also have *extrinsic scatter*, or *sample variance*, in both dimensions, i.e. scatter that is greater than can be accounted for by the error bars alone. Unlike statistical uncertainty, which is specified for each data point, extrinsic scatter is something that typically must be parameterized and fit to. Hence, it is part of the model. In this thesis, we colloquially refer to this extrinsic scatter as *slop*.

Slop, or sample variance, is a catch-all quantification of our ignorance of all "higher-order" physical processes that superimpose variations on the overall trend of the data and its intrinsic scatter, that we either can not or choose not to model explicitly. For example, in fitting models to GRB afterglows, we typically model the intrinsic flux density F_ν as a series of smoothly broken power laws in time. But there may be lower-level, shorter-timescale variability superimposed on these smooth light curves,

e.g., due to clumpiness in the circumburst medium or variability of the central engine. We typically model such variability to be normally distributed in $\log F_\nu$, and describe it with a single slop parameter, $\sigma_{\log F_\nu}$, which we fit to as a free model parameter. In the following formulation of the very general case of fitting models to data in two dimensions, we assume that slop is normally distributed in both dimensions, and can be described by two parameters (σ_x, σ_y) that are constant over the entire data set.[2]

In this section, we discuss three approaches for defining the likelihood function for this most general two-dimensional case: that of D'Agostini (2005, hereafter D05), that of Reichart (2001, hereafter R01), and a new approach, TRF (Trotter, Reichart & Foster 2011, in preparation), which is the main topic of this chapter of the thesis. All three statistics are based on the same fundamental formulation of the likelihood function in terms of the joint probability of the intrinsic and extrinsic probability distributions of the data set with the model, which we present below. Their differences we present in §2.3.

2.2.1 Analytical Approximation of the Likelihood Function

The following derivation follows closely that of R01, which formed the basis of much of the new investigation in this thesis. We wish to quantify the fitness of a model to a two-dimensional set of N measurements. The data points have centroids at $\{x_n, y_n\}$, and normally distributed intrinsic scatter, or error bars, $\{\sigma_{x,n}, \sigma_{y,n}\}$. We describe the model itself not as a curve, but rather as a *model distribution*, which we treat mathematically as a continuous distribution of points, or a relative probability distribution,

[2]Of course, more complicated parameterizations of slop are possible, including non-normal distributions and distributions that vary across the data set. For example, there may be a component of unmodeled variability in a GRB afterglow that is normally distributed in linear flux space, due to radio scintillation in the interstellar medium along the line of sight. In that case, our model would include an additional slop parameter, σ_{F_ν}, and the unmodeled lower-level, shorter-timescale variability would be described by a convolution of linear- and log-normal probability distributions. Such refinements are introduced as needed (see, e.g., §§2.7 & 3.4.2).

$g(x, y)$ along a curve $y_c(x; \vartheta_m)$, convolved with a two-dimensional Gaussian probability distribution function. The functional form of the curve is defined by a set of M parameters $\{\vartheta_m\}$. The width of the convolving Gaussian in each dimension is (σ_x, σ_y), and reflects the extrinsic scatter, or slop, in the data. In all that follows, we assume that slop in each dimension is constant across the entire data set. We also assume that both the intrinsic error bars and extrinsic slops in the x- and y-dimensions are uncorrelated.[3] Figure 2.4 illustrates a generic two-dimensional data set and model distribution. The inset box is expanded in Figure 2.5, which illustrates the overlapping intrinsic probability distribution of a single data point with the slop-convolved model probability distribution.

In order to proceed, the effectively one-dimensional model curve must be expressed in terms of a one-dimensional delta function. But at this point, we are completely free to choose the direction in which that delta function is defined. Most generally, the probability density along the curve can be expressed as $g(x, y)\delta(v_n - v_{c,n}(u_n; \vartheta_m))$. Here, the coordinates (u_n, v_n) are related to (x, y) by a simple rotation of the axes, and $v_{c,n}(u_n; \vartheta_m)$ describes the curve $y_c(x; \vartheta_m)$ in this rotated coordinate system. Note that, in general, the choice of (u_n, v_n) may be *different for each data point*. As we will see, the choice of rotated coordinates (u_n, v_n) is critical in determining the properties of the statistic we derive; the three statistics D05, R01 and TRF differ from each other only in the choice of (u_n, v_n). Convolving the probability density along the model curve with the two-dimensional Gaussian representing extrinsic scatter in the data, we obtain the *model distribution function*:

$$p_n^{\rm mod}(x', y'|\vartheta_m, \sigma_x, \sigma_y) = \int_{u_n} \int_{v_n} g(x, y)\delta(v_n - v_{c,n}(u_n; \vartheta_m))G(x', x, \sigma_x)G(y', y, \sigma_y)dv_n du_n \,, \tag{2.2}$$

[3]The treatment of correlated uncertainties in the data will be the subject of future research.

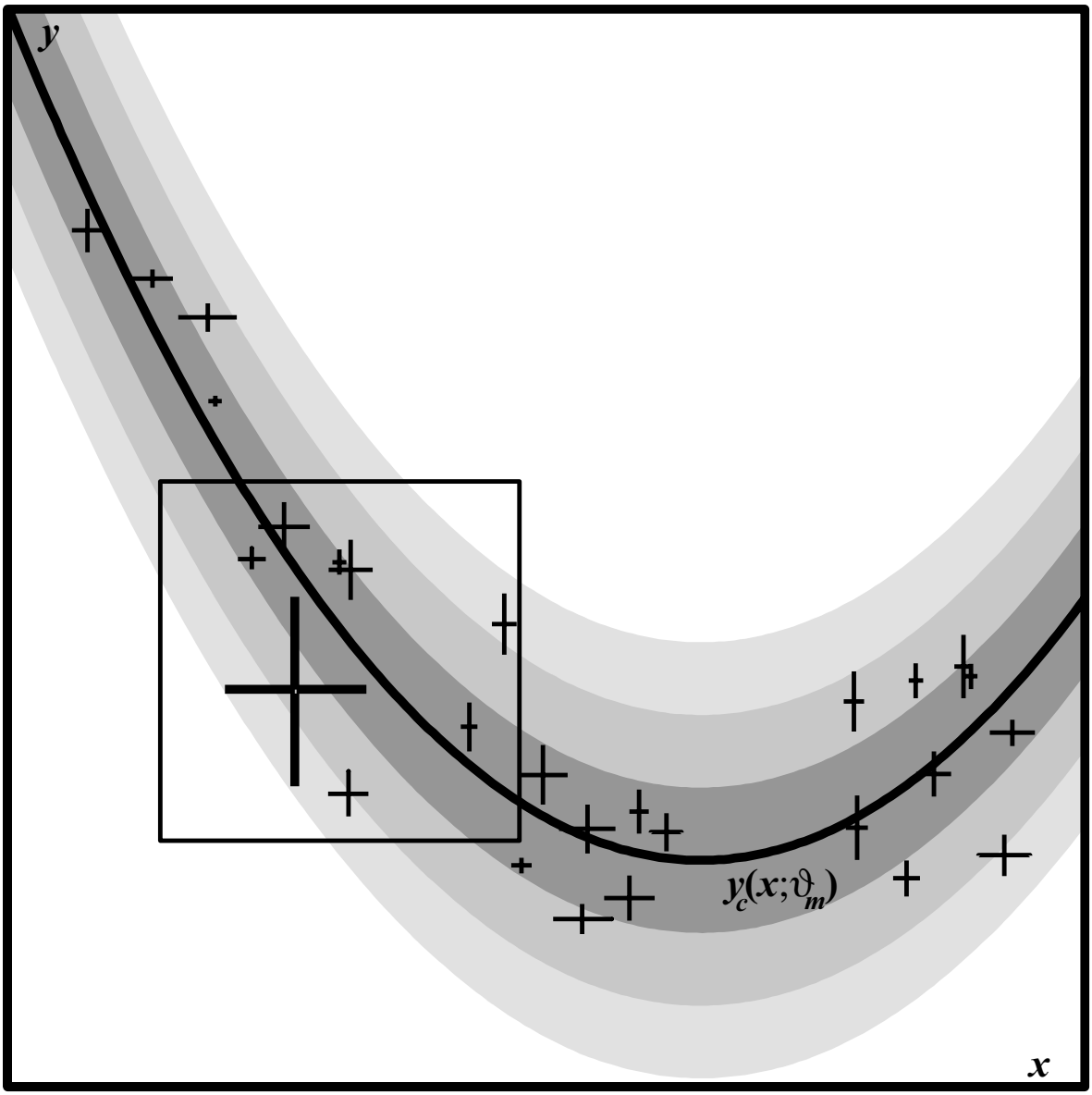

Figure 2.4 Illustration of a two-dimensional data set and model distribution. Inset box is expanded in Figure 2.5.

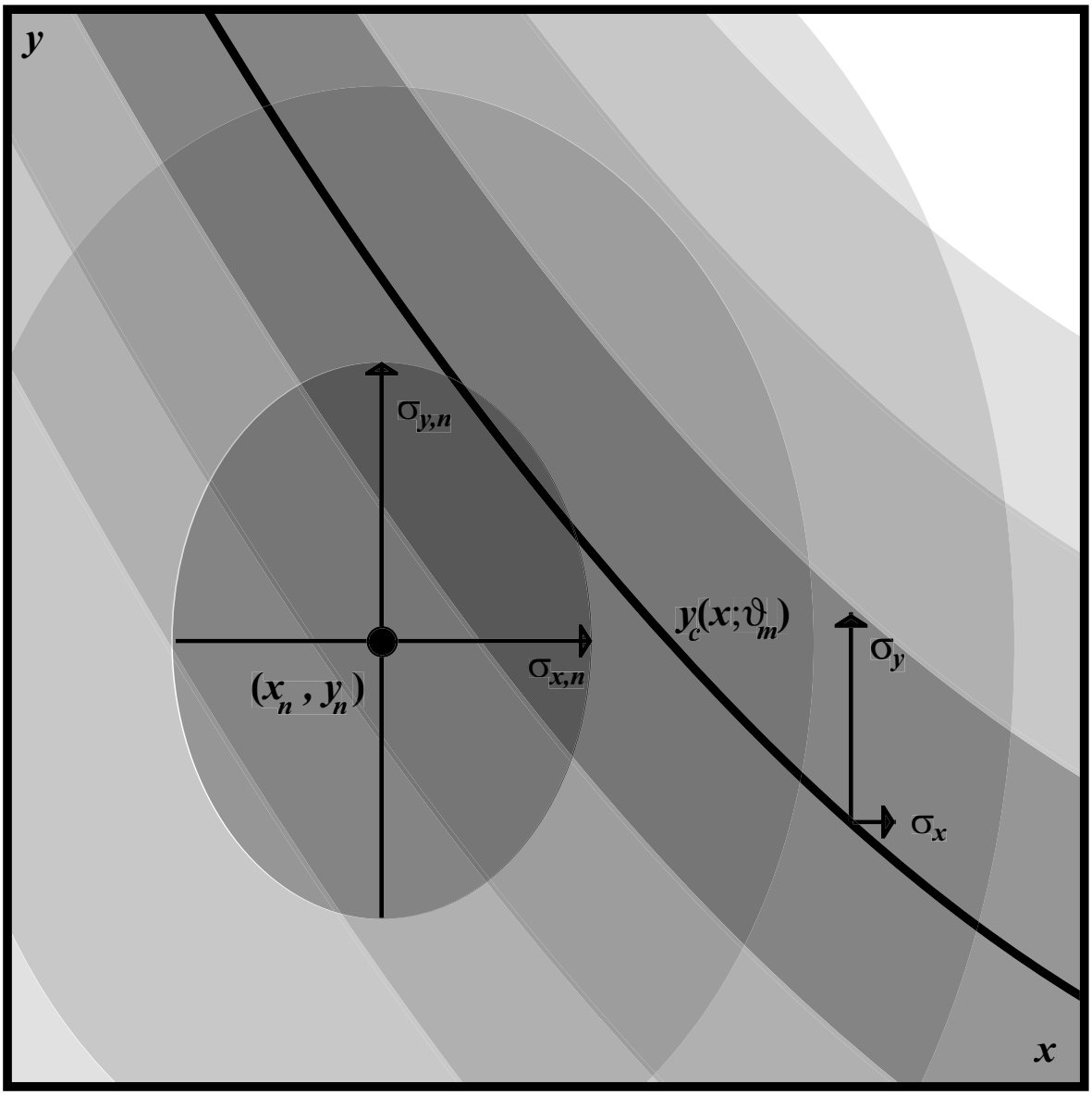

Figure 2.5 Expansion of the inset box of Figure 2.4, illustrating the intrinsic probability distribution of a single data point and the model distribution. The data point (x_n, y_n) has intrinsic error bars $(\sigma_{x,n}, \sigma_{y,n})$. The model distribution is the convolution of a probability density $g(x, y)$ along a curve $y_c(x; \vartheta_m)$ with a two-dimension Gaussian of width (σ_x, σ_y) that represents the extrinsic scatter, or slop, in the data set. Shading represents the 1, 2 and 3σ widths of the intrinsic probability distribution of the data point, and of the extrinsic slop-convolved model distribution.

where G denotes the normalized Gaussian function:

$$G(x', x, \sigma_x) \equiv \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2}\frac{(x'-x)^2}{\sigma_x^2}\right], \tag{2.3}$$

and where the integrals are all definite integrals from $-\infty$ to $+\infty$. For now, we continue to express the Gaussians, and $g(x, y)$ in terms of unrotated coordinates (x, y).

However, in reality, the model distribution is not what is observed. Let $f(x', y')$ represent the *data selection function*, which describes the efficiency at which the actual data set in question samples the model distribution. Then, the *observed* model distribution is:

$$p_n^{\rm obs}(x', y'|\vartheta_m, \sigma_x, \sigma_y) = \int_{u_n}\int_{v_n} f(x', y')g(x, y)\delta(v_n - v_{c,n}(u_n; \vartheta_m))\times \\ G(x', x, \sigma_x)G(y', y, \sigma_y)dv_n du_n\,. \tag{2.4}$$

We now consider the *joint probability* of a single data point with the observed model distribution of Equation 2.4. The intrinsic probability distribution of data point n is given by the two-dimensional Gaussian:

$$p_n^{\rm int}(x', y'|x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) = G_n(x', x_n, \sigma_{x,n})G_n(y', y_n, \sigma_{y,n})\,. \tag{2.5}$$

The joint probability of data point n with the model distribution is then obtained by integrating the product of $p_n^{\rm obs}$ and $p_n^{\rm int}$ over the (x', y') plane:

$$p_n(\vartheta_m, \sigma_x, \sigma_y|x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) = \\ \int_{x'}\int_{y'}\int_{u_n}\int_{v_n} f(x', y')g(x, y)\delta(v_n - v_{c,n}(u_n; \vartheta_m)) \times \\ G(x', x, \sigma_x)G(y', y, \sigma_y)G_n(x', x_n, \sigma_{x,n})G_n(y', y_n, \sigma_{y,n})dv_n du_n dy' dx'\,. \tag{2.6}$$

To achieve a computationally practical statistic, an analytic result is desirable. To this end, we now make a number of important simplifying assumptions. First, we assume that the selection function $f(x', y')$ varies slowly with respect to the scale of the intrinsic uncertainties $(\sigma_{x,n}, \sigma_{y,n})$ and extrinsic scatter (σ_x, σ_y), and so can be approximated as a constant in the (x', y') integral. In this case, the integral can be evaluated analytically, and the joint probability becomes:

$$p_n(\vartheta_m, \sigma_x, \sigma_y | x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) \approx \tag{2.7}$$
$$f(x_n, y_n) \int_{u_n} \int_{v_n} g(x, y) \delta(v_n - v_{c,n}(u_n; \vartheta_m)) G_n(x, x_n, \Sigma_{x,n}) G_n(y, y_n, \Sigma_{y,n}) dv_n du_n \, ,$$

where $(\Sigma_{x,n}, \Sigma_{y,n})$ are the quadrature sums of the intrinsic and extrinsic scatters:

$$\begin{aligned} \Sigma_{x,n} &\equiv \left(\sigma_{x,n}^2 + \sigma_x^2\right)^{1/2} \\ \Sigma_{y,n} &\equiv \left(\sigma_{y,n}^2 + \sigma_y^2\right)^{1/2} \, . \end{aligned} \tag{2.8}$$

In other words, the joint probability of data point n with the model distribution is proportional to the integral of the effectively one-dimensional probability density along the model curve through a two-dimensional *convolved* Gaussian whose widths are the quadrature sums of the intrinsic uncertainties and the slops in each direction (i.e., Figure 2.5 is equivalent to Figure 2.6).

We now make two further important simplifying assumptions: (1) the intrinsic probability density along the model curve $g(x, y)$ varies slowly with respect to the scale of the convolved error ellipse $(\Sigma_{x,n}, \Sigma_{y,n})$; and (2) the model curve $y_c(x; \vartheta_m)$ (or $v_{c,n}(u_n; \vartheta_m)$) deviates slowly from linearity over this same scale. The first assumption allows us to pull $g(x, y)$ out of the (u, v) integral, because it is effectively a constant across the integral. The second assumption requires some finesse: We approximate the model curve $y_c(x; \vartheta_m)$ as a line $y_{t,n}(x)$ passing through the point $(x_{t,n}, y_{t,n})$ where the

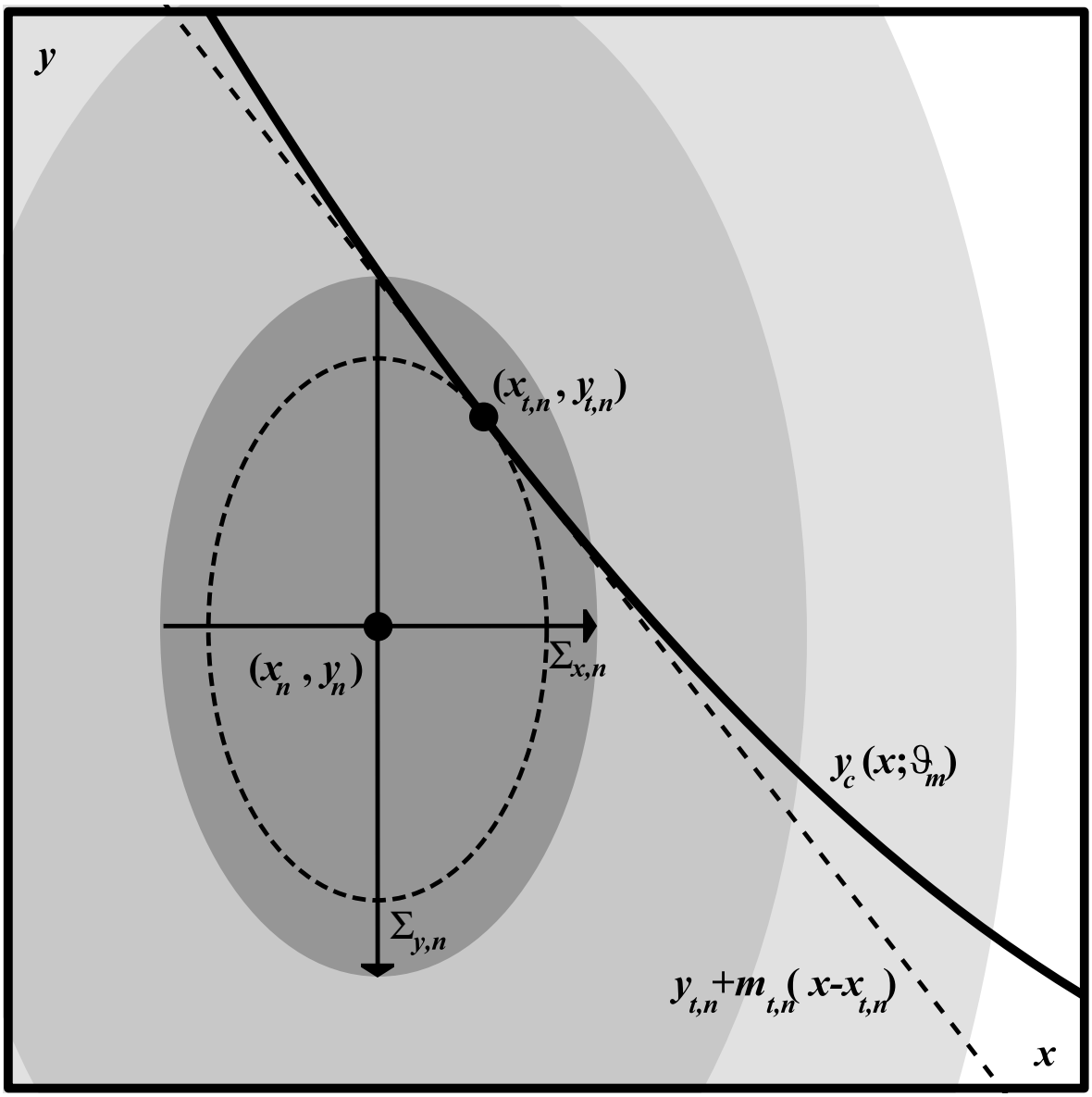

Figure 2.6 Illustration of the convolved error ellipse and tangent line approximation. The data point (x_n, y_n) is assigned convolved error bars $(\Sigma_{x,n}, \Sigma_{y,n})$. The model distribution is replaced by the curve $y_c(x)$. For non-linear models, this curve is approximated as a line $y_{t,n}(x)$ tangent to the convolved error ellipse at point $(x_{t,n}, y_{t,n})$.

curve is tangent to the convolved error ellipse, with a slope $m_{t,n}$ equal to the slope of the curve at that tangent point (Figure 2.6):

$$y_c(x) \approx y_{t,n} + m_{t,n}(x - x_{t,n})\,. \tag{2.9}$$

To find the tangent point, consider an arbitrary point x along the model curve $y_c(x;\vartheta_m)$. If the curve at this point is a tangent point $(x_{t,n}, y_{t,n})$ to the error ellipse, then:

$$\frac{(y_c(x;\vartheta_m) - y_n)^2}{\Sigma_{y,n}^2} + \frac{(x - x_n)^2}{\Sigma_{x,n}^2} = \frac{(y_{t,n} - y_n)^2}{\Sigma_{y,n}^2} + \frac{(x_{t,n} - x_n)^2}{\Sigma_{x,n}^2}\,. \tag{2.10}$$

Note that the right-hand side of this equation is a constant. If we take the derivative of Equation 2.10, we find that the condition for $(x, y_c(x;\vartheta_m))$ to be a tangent point is equivalent to requiring:

$$\frac{d}{dx}\left(\frac{(y_c(x;\vartheta_m) - y_n)^2}{\Sigma_{y,n}^2} + \frac{(x - x_n)^2}{\Sigma_{x,n}^2}\right) = 0\,. \tag{2.11}$$

In other words, the tangent point is equivalent to the point on the curve $y_c(x;\vartheta_m)$ that minimizes the radial distance to the error ellipse centroid, as measured in units of $(\Sigma_{x,n}, \Sigma_{y,n})$. Carrying out the derivative, we see that the tangent point $x_{t,n}$ is the value of x for which:

$$(y_c(x) - y_n)\frac{dy_c(x;\vartheta_m)}{dx}\Sigma_{x,n}^2 + (x - x_n)\Sigma_{y,n}^2 = 0\,. \tag{2.12}$$

For a linear model, i.e., $y_c(x;\vartheta_m) = mx + b$, the tangent point coordinates can be obtained from Equation 2.12 analytically:

$$\begin{aligned} x_{t,n} &= \frac{m\Sigma_{x,n}^2\,(y_n - b) + \Sigma_{y,n}^2 x_n}{m^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\ y_{t,n} &= mx_{t,n} + b\,. \end{aligned} \tag{2.13}$$

For non-linear model curves, the tangent point must be found numerically. Depending on the functional form of $y_c(x;\vartheta_m)$, this point can be bracketed and found by any number of numerical root-finding algorithms.[4] We note that in the case of non-monotonic curves, there may be two or more such tangent points for each data point; in those cases, we choose the one that maximizes the joint probability.

Finally, under all these simplifying assumptions, the joint probability of data point n with the model distribution of Equation 2.7 can be evaluated analytically by integrating over v_n:

$$\begin{aligned} & p_n(\vartheta_m, \sigma_x, \sigma_y | x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) \\ \approx \quad & f(x_n, y_n) g(x_n, y_n) \int_{-\infty}^{\infty} G_n(x, x_n, \Sigma_{x,n}) G_n(y_c(x;\vartheta_m), y_n, \Sigma_{y,n}) du_n \, . \end{aligned} \tag{2.14}$$

Re-expressing $du_n = \frac{du_n}{dx} dx$, substituting $y_c(x;\vartheta_m) \approx y_{t,n} + m_{t,n}(x - x_{t,n})$, and integrating over x, we obtain:

$$\begin{aligned} & p_n(\vartheta_m, \sigma_x, \sigma_y | x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) \\ \approx \quad & f(x_n, y_n) g(x_n, y_n) \frac{du_n}{dx} G_n \left[y_n, y_{t,n} + m_{t,n}(x_n - x_{t,n}), \sqrt{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right] \, . \end{aligned} \tag{2.15}$$

As a reminder, (u_n, v_n) is the rotated coordinate system in which we define the fundamentally one-dimensional curve $\delta(v_n - v_{c,n}(u_n;\vartheta_m))$ that underlies the observed model distribution (Equation 2.2). In this derivation of the joint probability of a data point with the model distribution, there is no a priori reason that one choice of u_n should be preferable to another, or even that u_n should be the same for all data points. But, as we shall see in §2.4, *different choices of* $\mathrm{u_n}$ *yield different statistics*, entirely due to the factor $\frac{du_n}{dx}$ in the joint probability. And each resulting statistic exhibits different

[4] R01 presents an alternative, iterative algorithm for determining the tangent point. We have found that this algorithm, though efficient when it succeeds, often fails to converge, especially for non-monotonic model curves.

behaviors and yields different best-fit model distributions for a given data set.

The likelihood function $\mathcal{L}$ is defined as the joint probability of the model distribution with all the data points, and is simply the product of the individual joint probabilities:

$$\mathcal{L} \equiv \prod_{n=1}^{N} p_n(\vartheta_m, \sigma_x, \sigma_y | x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) \,. \tag{2.16}$$

In practice, we find the best-fit model parameters by minimizing $-2\ln\mathcal{L}$:

$$\begin{aligned} -2\ln\mathcal{L} &= -2\sum_{n=1}^{N} \ln p_n(\vartheta_m, \sigma_x, \sigma_y | x_n, y_n, \sigma_{x,n}, \sigma_{y,n}) \\ &= \sum_{n=1}^{N} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})\right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\ &\quad -2\sum_{n=1}^{N} \ln\left(\frac{du_n}{dx} \frac{1}{\sqrt{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}}\right) + C \,, \end{aligned} \tag{2.17}$$

where C is an (undetermined) constant for a given data set:

$$C = N\ln(2\pi) - 2\sum_{n=1}^{N} \ln f(x_n, y_n) g(x_n, y_n) \,. \tag{2.18}$$

We choose to define the fitness in terms of $-2\ln\mathcal{L}$ because it is analogous to χ^2 in certain simplifying cases. For instance, if there are no error bars in the x-direction, no slop in either direction, and $\frac{du_n}{dx} = 1$ for all points[5], then:

$$\begin{aligned} -2\ln\mathcal{L} &= \sum_{n=1}^{N} \frac{\left[y_c(x_n) - y_n\right]^2}{\sigma_{y,n}^2} + C \\ &\equiv \chi^2 + C \,. \end{aligned} \tag{2.19}$$

In *any* case, the additive constant C in this formulation does not affect the relative

[5]As we will show in §§2.3 & 2.4.2, $\frac{du_n}{dx} = 1$ for all points for the D05 statistic, and $\frac{du_n}{dx} \to 1$ for the TRF statistic in this one-dimensional limiting case of no error bars or slop in the x-direction.

fitness $-2\ln\mathcal{L} - (-2\ln\mathcal{L})_{\min}$ of a model with respect to the best fit; in this simple one-dimensional case, the relative fitness reduces to $\Delta\chi^2$.

2.2.2 Summary of Simplifying Assumptions

Before proceeding to discuss the different statistics that result from different choices of u_n, we summarize the simplifying assumptions that are common to all three in the analysis so far:

- Intrinsic scatter (error bars) in both dimensions is uncorrelated and normally distributed.
- Extrinsic scatter (slop) in both dimensions is uncorrelated and normally distributed.
- The selection function $f(x, y)$ and model probability density $g(x, y)$ vary slowly with respect to both the intrinsic and extrinsic scatter, and so can be approximated as constants over these scales.
- Non-linear model curves vary slowly with respect to the intrinsic and extrinsic scatter, and so can be approximated by lines tangent to convolved error ellipses over these scales.

Finally, consider the example of Figure 2.7. It is worth noting that although the large error bars of the data point in bold violate the latter two simplifying assumptions, its ability to help constrain the model parameters is clearly negligible, compared to the other data points. Consequently, in such cases the statistics we present below can still be taken as valid.

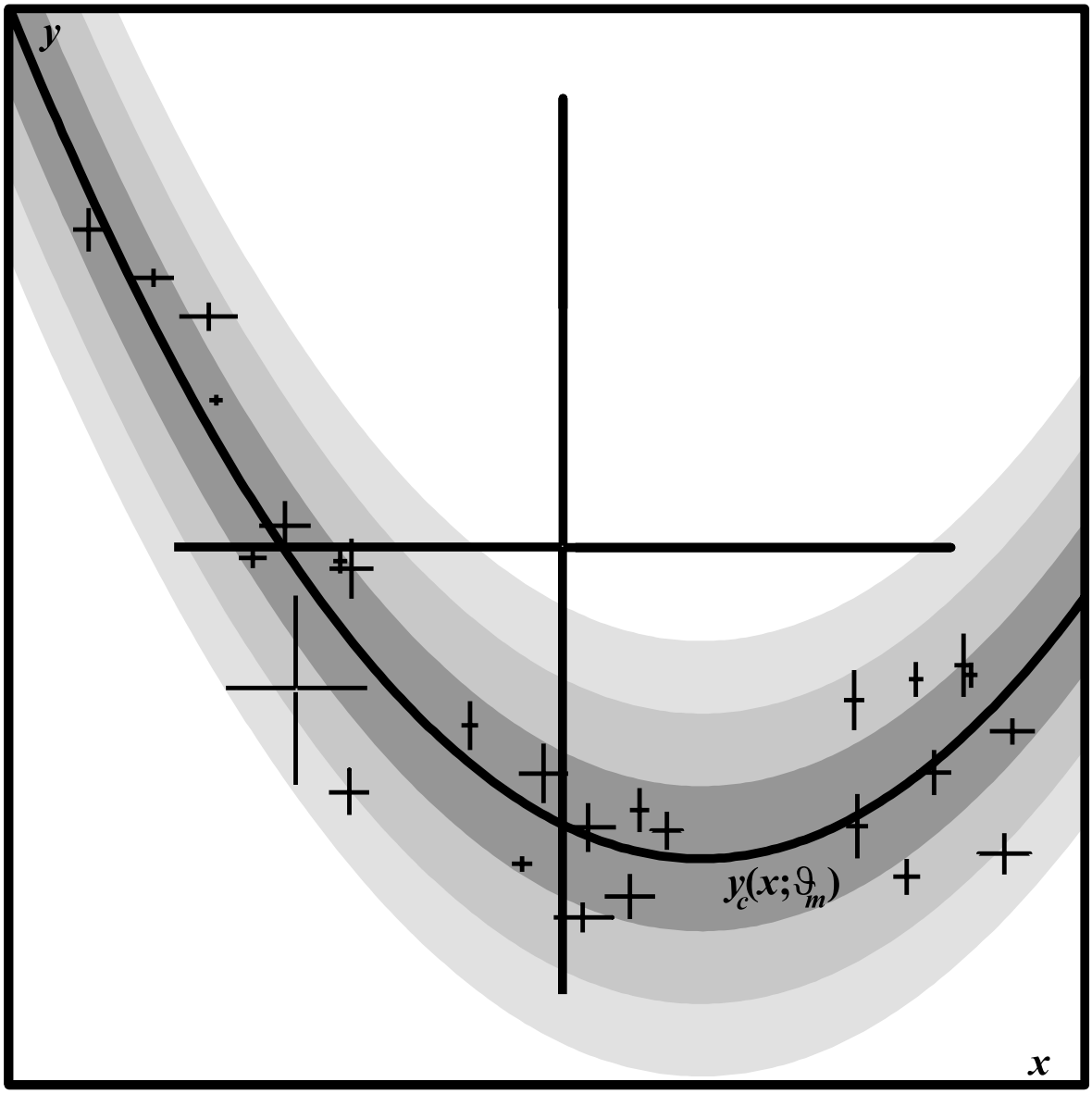

Figure 2.7 Illustration of a two-dimensional data set and model distribution, with a single anomalous data point (in bold).

2.3 Three Statistics: D05, R01 and TRF

As hinted at above, the choice of path integral element du_n is, in fact, the only choice that differentiates the three statistics we explore in this thesis. As we will see, each statistic has its own strengths and its own weaknesses. There may well be no two-dimensional statistic of this probabilistic form that is ideal for all model fitting cases. In this section, we describe the basic properties, including the functional form of the likelihood function, of three statistics. In §2.4 we discuss, compare and contrast their behavior in some detail for various special cases and coordinate transformations.

D05 presents a statistic for which du_n is simply dx, so that $\frac{du_n}{dx} = 1$ for all data points. In this case, the likelihood has the form:

$$\begin{aligned}
\mathcal{L}^{\rm D05} &\propto \prod_{n=1}^{N} \frac{1}{\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{-\frac{1}{2}\frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})\right]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right\} \\
-2\ln\mathcal{L}^{\rm D05} &= \sum_{n=1}^{N} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})\right]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\
&\quad + \sum_{n=1}^{N} \ln(m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2) + C\,. \qquad (2.20)
\end{aligned}$$

The D05 statistic can be seen to be analogous to a one-dimensional χ^2 statistic in y, where the difference between the model and the data point is the difference between the tangent line at $x = x_n$ and y_n, and where the 1σ uncertainty in the convolved data point is replaced by the quadrature sum of $\Sigma_{y,n}$ and $\Sigma_{x,n}$ projected into the y-direction using the slope $m_{t,n}$. D05 differs from a traditional χ^2 statistic in that the denominator of the argument of the exponential, and the prefactor of the exponential are themselves functions of the slops (σ_x, σ_y), which are treated as free model parameters.

It should be noted that a common approach to fitting models to data with error bars in both dimensions, and with unmodeled extrinsic scatter, is to use a one-dimensional χ^2

statistic in y, with the intrinsic $\sigma_{x,n}$ error bars projected into the y-direction, as here, but to ignore the prefactor and instead approximate the slop either by adding an additional uncertainty σ_y to the intrinsic error bars in quadrature, or even by multiplying the intrinsic error bars by some constant, and iterating until the reduced $\chi^2 = 1$. In general, however, this is not correct, even in the one-dimensional limiting case where error bars and slop are entirely in the y-direction. It amounts to holding σ_y fixed at its best-fit value, and, much as holding $H_0 = 72$ or $R_V^{\rm MW} = 3.1$ in the examples presented in §2.1, it is unnecessarily restrictive. Rather, since the exponential term in the likelihood of Equation 2.20 decreases as the slops go to zero, while the prefactor term decreases as the slops go to infinity, maximizing the likelihood with the slops as free parameters naturally constrains (σ_x, σ_y), without unnecessarily restricting them. In the one-dimensional limit, we refer to this more proper approach as a "χ^2-like" statistic.

As we will discuss in §2.4.1, D05 is *non-invertible*: a fit to x vs. y will, in general, yield a different model curve than a fit to y vs. x. R01 presented a first attempt at an *invertible statistic*, that is, a statistic that is invariant under an exchange of the x- and y-axes. The R01 statistic is formulated using $du_n = ds_n$, where $ds_n = (dx^2 + dy^2)^{1/2}$ is parallel to the tangent line, with slope $m_{t,n}$. The net effect is to add a multiplicative prefactor $\frac{du_n}{dx} = \sqrt{1 + m_{t,n}^2}$ to the joint probability:

$$\begin{aligned}
\mathcal{L}^{\rm R01} &\propto \prod_{n=1}^{N} \sqrt{\frac{1 + m_{t,n}^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{ -\frac{1}{2} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\} \\
-2 \ln \mathcal{L}^{\rm R01} &= \sum_{n=1}^{N} \frac{\left[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n}) \right]^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\
&\quad - \sum_{n=1}^{N} \ln \left(\frac{1 + m_{t,n}^2}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right) + C \,. \qquad (2.21)
\end{aligned}$$

It can be easily shown that $\mathcal{L}^{\rm R01}$ is analytically invariant under inversion of the x- and

y-axes (see §2.4.1). However, R01 does not reduce to a one-dimensional χ^2-like statistic in the limit $\Sigma_{x,n} \to 0$ (see §2.4.2). One can also show that R01 is not invariant under multiplicative scaling transformations of either axis (see §2.4.4). The discovery of these limitations in R01 prompted us to explore other possible forms of the likelihood that preserve invertibility.

We present a new statistic, TRF, that is both invertible and reduces to a χ^2-like statistic in the one-dimensional limit. We propose the following functional form of the likelihood:

$$\begin{aligned}
\mathcal{L}^{\mathrm{TRF}} &\propto \prod_{n=1}^{N} \sqrt{\frac{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}} \exp\left\{-\frac{1}{2}\frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\right\} \\
-2\ln\mathcal{L}^{\mathrm{TRF}} &= \sum_{n=1}^{N} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} \\
&\quad - \sum_{n=1}^{N} \ln\left(\frac{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}\right) + C\,. \qquad (2.22)
\end{aligned}$$

Geometrically, TRF is equivalent to choosing du_n to be perpendicular to the line segment connecting the data point centroid (x_n, y_n) and the tangent point of the curve $(x_{t,n}, y_{t,n})$ (that is, the delta function $\delta(v_n - v_{c,n}(u_n))$ is defined along the axis that connects the centroid and the tangent point). As with R01, it can easily be shown that this form of $\mathcal{L}^{\mathrm{TRF}}$ is analytically invertible (see §2.4.1). TRF reduces to a χ^2-like one-dimensional statistic in both limits $\Sigma_{x,n} \to 0$ (in which case it is equivalent to D05) and $\Sigma_{y,n} \to 0$ (see §2.4.2). Furthermore, in the case $\Sigma_{x,n} = \Sigma_{y,n}$, TRF is equivalent to R01. Most fortuitously, we find that *TRF is analogous to a one-dimensional χ^2-like statistic measured in the direction of the tangent point*, which we will now derive.

Figure 2.8 illustrates the geometry of the TRF statistic. The convolved error ellipse of a given point has a centroid at point $O \equiv (x_n, y_n)$, with 1σ widths $(\Sigma_{x,n}, \Sigma_{y,n})$. The tangent point of the model curve $y_c(x; \vartheta_m)$ is at point $T \equiv (x_{t,n}, y_{t,n})$. As described

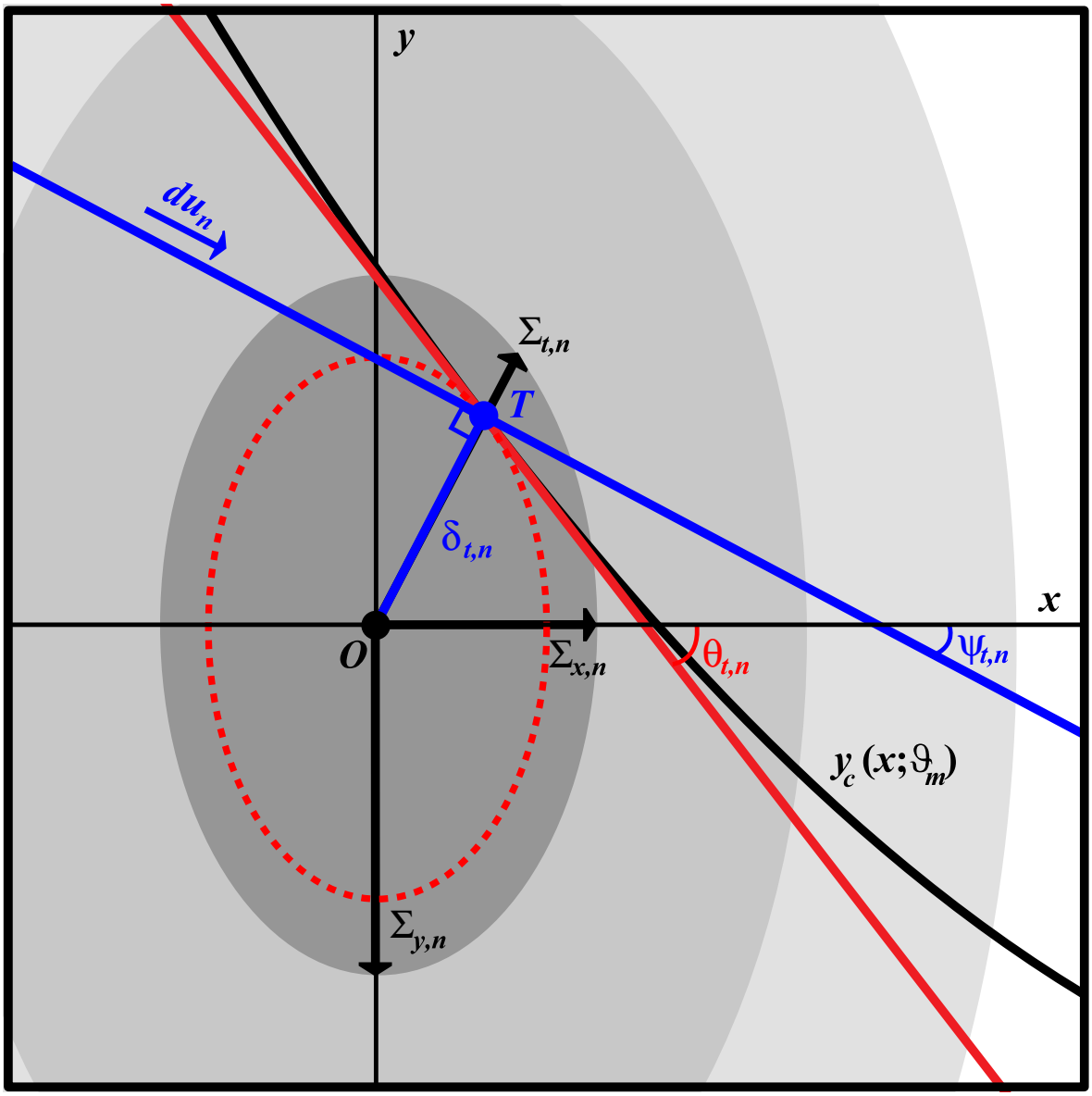

Figure 2.8 Illustration of the geometry of the TRF statistic. Data point centroid is at (x_n, y_n) (point O), with convolved widths $(\Sigma_{x,n}, \Sigma_{y,n})$. Model curve $y_c(x;\vartheta_m)$ is tangent to the convolved error ellipse at tangent point $(x_{t,n}, y_{t,n})$ (point T). Red line is the linear approximation of the model curve, with slope $m_{t,n} = \tan\theta_{t,n}$. Blue line indicates the direction du_n of path integration for the TRF statistic, perpendicular to the segment $\overline{OT}$. TRF is geometrically equivalent to a 1D χ^2-like statistic measured in the direction of the tangent point (see Equation 2.28).

above, we approximate the curve $y_c(x;\vartheta_m)$ as a line through that point, with slope $m_{t,n} = \tan\theta_{t,n}$.

For TRF, we define the rotated coordinate system (u_n, v_n) such that u_n is perpendicular to the segment connecting the ellipse centroid and the tangent point. The u_n-axis is rotated by an angle $\psi_{t,n}$ with respect to the x-axis, and so differs from D05. The u_n-axis describes an angle $\psi_{t,n} - \theta_{t,n}$ with respect to the tangent line, and so differs from R01.

It is not difficult to show that the radial distance of the tangent point from the centroid of the ellipse is:

$$\delta_{t,n} \equiv \overline{OT} = [y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})] \frac{\sqrt{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}\,, \tag{2.23}$$

and that the 1σ radius of the convolved error ellipse in the direction of the tangent point is:

$$\Sigma_{t,n} = \frac{\Sigma_{x,n}\Sigma_{y,n}}{\sqrt{\Sigma_{x,n}^2\cos^2\psi_{t,n} + \Sigma_{y,n}^2\sin^2\psi_{t,n}}}\,, \tag{2.24}$$

where

$$\begin{aligned} \cos^2\psi_{t,n} &= \left(\frac{y_{t,n} - y_n}{\delta_{t,n}}\right)^2 \\ \sin^2\psi_{t,n} &= \left(\frac{x_{t,n} - x_n}{\delta_{t,n}}\right)^2\,. \end{aligned} \tag{2.25}$$

Substituting for $(x_{t,n}, y_{t,n})$ from Equation 2.13, we find that:

$$\Sigma_{t,n} = \sqrt{\frac{m_{t,n}^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}}\,, \tag{2.26}$$

and

$$\frac{\delta_{t,n}}{\Sigma_{t,n}} = \frac{y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})}{\sqrt{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \,. \tag{2.27}$$

Thus, we see that the expression for $\mathcal{L}^{\mathrm{TRF}}$ in Equation 2.22 can equivalently be expressed as:

$$\begin{aligned} \mathcal{L}^{\mathrm{TRF}} &\propto \prod_{n=1}^{N} \frac{1}{\Sigma_{t,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\Sigma_{t,n}}\right)^2\right] \\ -2\ln \mathcal{L}^{\mathrm{TRF}} &= \sum_{n=1}^{N} \frac{\delta_{t,n}^2}{\Sigma_{t,n}^2} + 2\sum_{n=1}^{N} \ln \Sigma_{t,n} + C \,. \end{aligned} \tag{2.28}$$

In other words, *TRF is analogous to a one-dimensional χ^2-like statistic, measured in the direction of the tangent point to the convolved error ellipse of each data point.* For future reference, we can also express the D05 likelihood of Equation 2.20 in terms of $\Sigma_{t,n}$; note that the argument of the exponential is equivalent; however, the prefactor differs considerably from the intuitive form of Equation 2.28:

$$\mathcal{L}^{\mathrm{D05}} \propto \prod_{n=1}^{N} \left(\frac{\sqrt{m_{t,n}^2 \Sigma_{x,n}^4 + \Sigma_{y,n}^4}}{m_{t,n}^2 \Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right) \frac{1}{\Sigma_{t,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\Sigma_{t,n}}\right)^2\right] . \tag{2.29}$$

2.4 Comparison of the D05, R01 and TRF Statistics

As a reminder, the different statistics that we are considering here differ only in the choice of the rotated coordinate system (u_n, v_n) in which we define the delta function that traces the model curve $y_c(x; \vartheta_m)$, through the factor $\frac{du_n}{dx}$ in the likelihood function (Equation 2.17). In this section, we consider the choices of $\frac{du_n}{dx}$ that yield the D05, R01 and TRF statistics (Equations 2.20, 2.21 & 2.22), and evaluate their likelihood functions for: invertibility (§2.4.1); behavior in one-dimensional limiting cases (§2.4.2); the "two-point test" (§2.4.3); scalability (§2.4.4); and degeneracy of slop in linear fits

(§2.4.5). We summarize the strengths and weaknesses of the three statistics in Table 2.1 (§2.4.6).

2.4.1 Invertibility: The Motivation for R01 and TRF

One desirable property of a 2D statistic is that it be *invertible*; that is, if the best-fit model curve to y vs. x is $y_c(x)$, the best-fit model curve to x vs. y should be $x_c(y) = y_c^{-1}(x)$. In the case of a linear fit, if the slope of the best-fit line to y vs. x is m_{xy}, a fit to x vs. y with an invertible statistic should give a best-fit slope $m_{yx} = 1/m_{xy}$. For non-linear model curves, $x_c(y)$ may or may not be an analytic function, and may have to be defined piece-wise or numerically; still, we expect, for an invertible statistic, that the slopes of the curves at each tangent point $m_{t,n} \to 1/m_{t,n}$ in the inverted space.

Most generally, for an invertible statistic, the likelihood should be analytically invariant under the substitutions $x_n \leftrightarrow y_n$, $x_{t,n} \leftrightarrow y_{t,n}$, $\Sigma_{x,n} \leftrightarrow \Sigma_{y,n}$, and $m_{t,n} \leftrightarrow 1/m_{t,n}$. If we perform these substitutions for the D05 joint probability (Equation 2.20), we find the inverted D05* likelihood:

$$\begin{aligned}
\mathcal{L}^{\mathrm{D05}*} &\propto \prod_{n=1}^{N} \frac{1}{\sqrt{\Sigma_{y,n}^2/m_{t,n}^2 + \Sigma_{x,n}^2}} \exp\left\{ -\frac{1}{2} \frac{[x_n - x_{t,n} - (y_n - y_{t,n})/m_{t,n}]^2}{\Sigma_{y,n}^2/m_{t,n}^2 + \Sigma_{x,n}^2} \right\} \\
&= \prod_{n=1}^{N} \frac{m_{t,n}}{\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{ -\frac{1}{2} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\} \\
&= \left(\prod_{n=1}^{N} m_{t,n} \right) \mathcal{L}^{\mathrm{D05}}, \qquad (2.30)
\end{aligned}$$

that is, $\mathcal{L}^{\mathrm{D05}*} \neq \mathcal{L}^{\mathrm{D05}}$. Note that in this form of the inverted D05* likelihood, all quantities are still defined in the original, uninverted y vs. x space; the best-fit curve obtained by maximizing $\mathcal{L}^{\mathrm{D05}*}$ should be plotted in the original space. Note also that the argument of the exponential, which is the same for all three statistics (D05, R01 and

TRF), is analytically invertible under these substitutions; the non-invertibility of D05 resides in the prefactor. However, it can be easily shown that the prefactors of R01 and TRF (Equations 2.21 & 2.22) are analytically invertible, so that inverted $\mathcal{L}^{\mathrm{R01}*} = \mathcal{L}^{\mathrm{R01}}$ and inverted $\mathcal{L}^{\mathrm{TRF}*} = \mathcal{L}^{\mathrm{TRF}}$.

2.4.1.1 Demonstration of Invertibility and D05 Bias: Gaussian Random Clouds

Consider the following simple, but illuminating, example: linear fits to circularly symmetric, slop-dominated, Gaussian random clouds of points. We generated ensembles of zero-mean, symmetric $\sigma_x = \sigma_y \equiv \sigma = 1$ clouds of $N = 100$ points each. The independent Gaussian random coordinates (x_n, y_n) of each point in each cloud were generated via the Box-Muller transformation on independent, uniformly distributed random numbers in the range $[0, 1]$. The data points were assigned zero intrinsic error bars $\sigma_{x,n} = \sigma_{y,n} = 0$.

In this simple example, since there is no intrinsic uncertainty in the data, $\Sigma_{x,n} = \sigma_x = \sigma$ and $\Sigma_{y,n} = \sigma_y = \sigma$ for all points, where (σ_x, σ_y) are the values of the fitted slop parameters. Since we are fitting a linear model $y_c(x) = b + mx$, the tangent point slope for all points is simply equal to the slope of the fitted line, $m_{t,n} = m \equiv \tan\theta$. Hence, the TRF likelihood of Equation 2.28 reduces to:

$$\mathcal{L}^{\mathrm{TRF}} \propto \prod_{n=1}^{N} \frac{1}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\sigma}\right)^2\right], \tag{2.31}$$

while the D05 likelihood of Equation 2.29 reduces to:

$$\begin{aligned}\mathcal{L}^{\mathrm{D05}} &\propto \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{1+m^2}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{t,n}}{\sigma}\right)^2\right] \\ &= \left(\frac{1}{\sqrt{1+m^2}}\right)^N \mathcal{L}^{\mathrm{TRF}} \\ &\equiv \left(\cos^N\theta\right)\mathcal{L}^{\mathrm{TRF}}. \end{aligned} \tag{2.32}$$

Note, again, that the argument of the exponential is always the same for all of the statistics we are considering; any difference in their predictions lies entirely in the prefactor. What Equation 2.32 tells us is that, for a given guess of a best-fit linear slope $m \equiv \tan\theta$, the likelihood of D05 differs from TRF by a factor $\cos^N\theta$. Since any model-fitting algorithm we might choose is solely concerned with maximizing the likelihood function (or, equivalently, with minimizing $-2\ln\mathcal{L}$), it should be clear that, in this simple example, D05 will attempt to maximize $\cos\theta$, and hence minimize the slope, so that $\theta \sim 0$, to the degree that the particular data set allows. Under TRF, however, there is no a priori bias towards $\theta = 0$.

TRF is analytically invertible, but D05 is not. If we perform the above analysis using the inverted D05* likelihood of Equation 2.30, we see that $\mathcal{L}^{\mathrm{D05*}} = m^N\mathcal{L}^{\mathrm{D05}}$, so that $\mathcal{L}^{\mathrm{D05*}} = \left(\sin^N\theta\right)\mathcal{L}^{\mathrm{TRF}}$. Fits using the inverted D05* statistic will thus be biased to $\theta = 90°$, as plotted in the original, uninverted space. TRF *will* choose a particular best-fit θ, depending on the particular accidental correlations in a given random cloud (no given cloud with a finite number of data points is truly perfectly symmetric); this slope will be the same under inversion (see Figure 2.9). However, if we conduct fits to an ensemble of many such Gaussian random clouds, we should expect to find a uniform probability distribution of fitted θ for TRF, while for D05, the probability distribution should be proportional to $\cos^N\theta$; we find that this is indeed the case (see Figure 2.10).

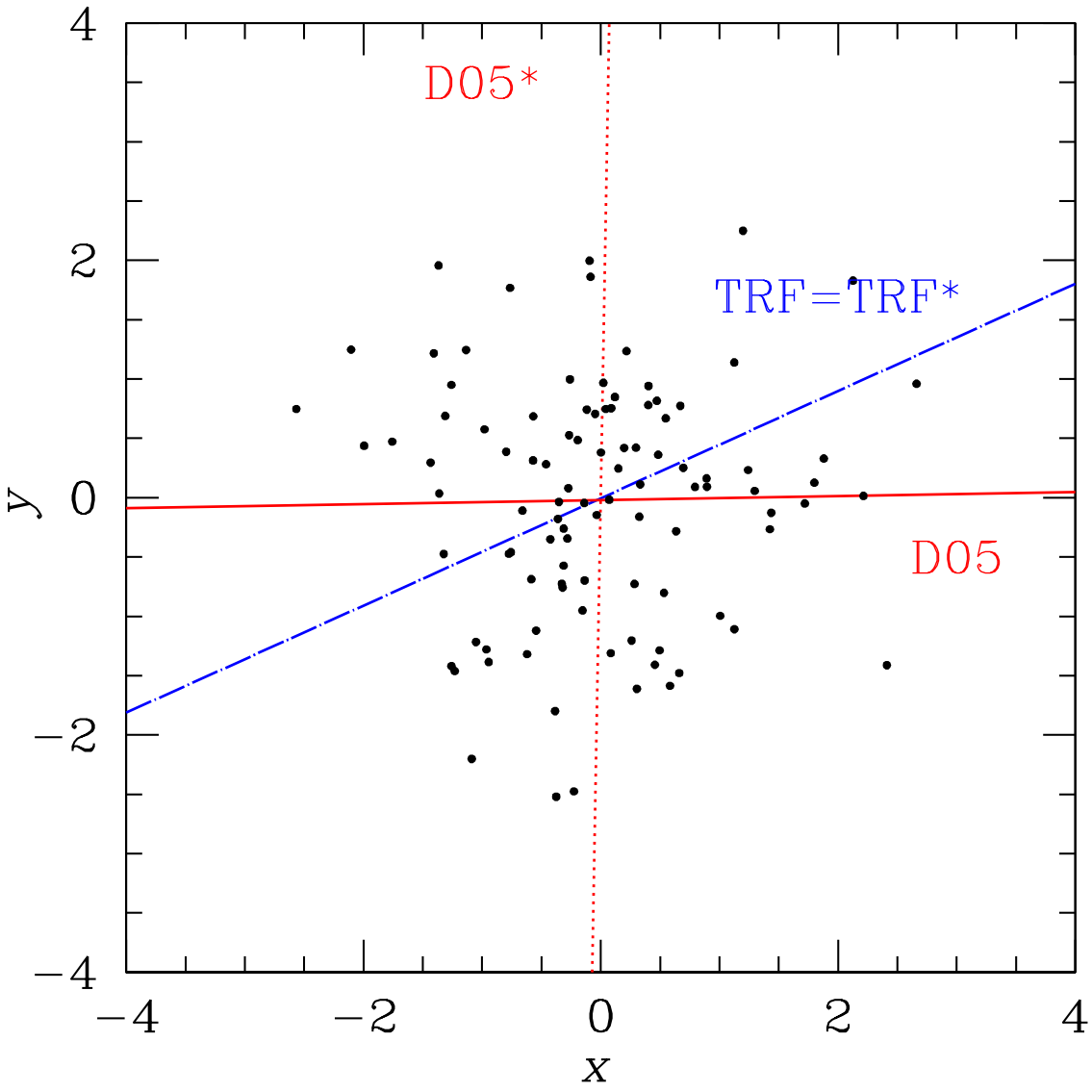

Figure 2.9 Linear fits to a slop-dominated Gaussian random cloud of $N = 100$ points with zero intrinsic error bars. Point coordinates generated with zero mean, $\sigma = 1$ Gaussian random variables in both dimensions. Solid red line shows fit with D05 statistic; dotted red line is fit with inverted D05* statistic. Blue line shows fit and inverted fit with TRF statistic; for an invertible statistic, the two are the same.

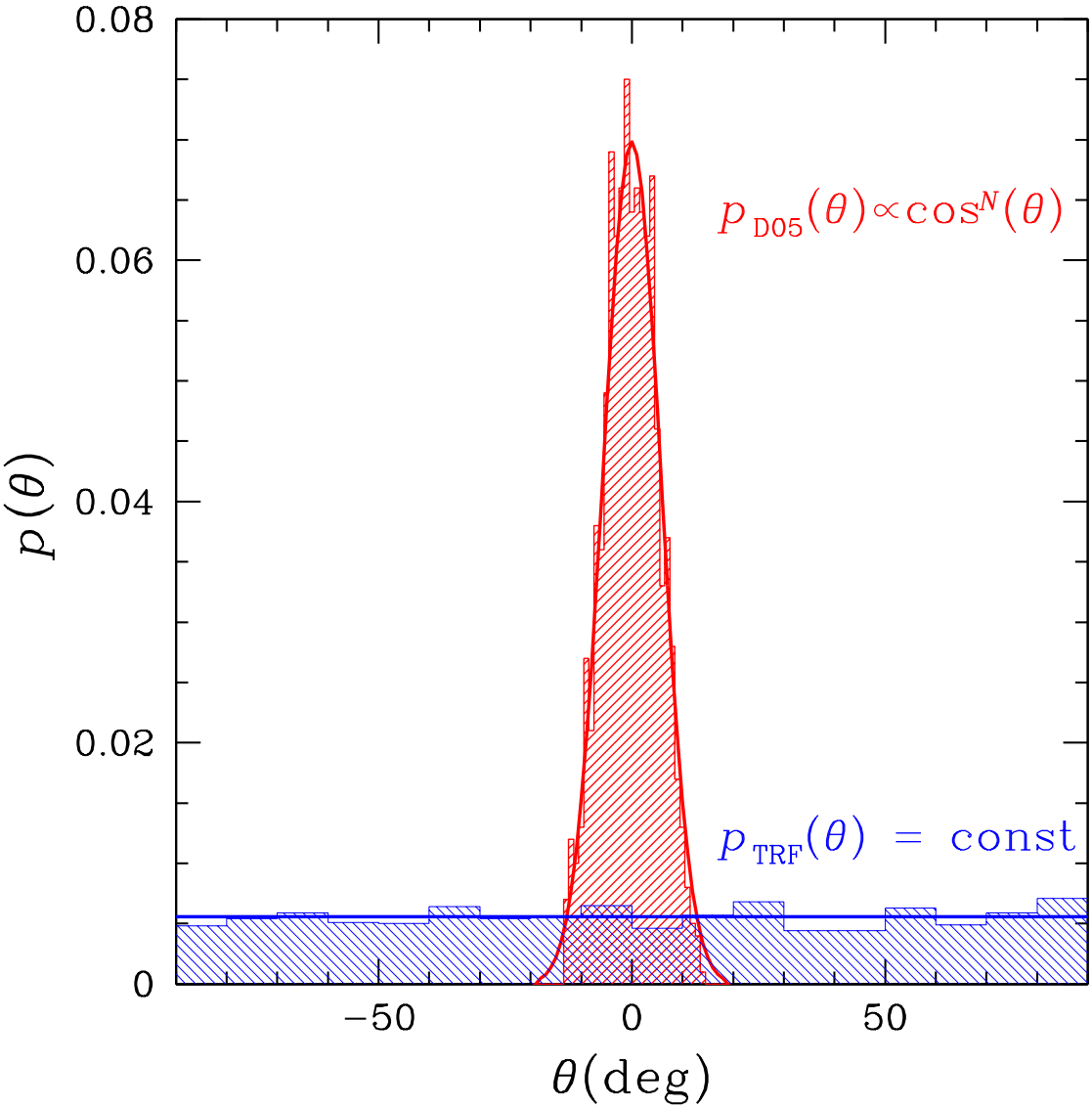

Figure 2.10 Histogram of fitted slope θ for an ensemble of 1000 zero-mean, $\sigma = 1$ Gaussian random clouds of $N = 100$ points each, for both D05 and TRF statistics. Solid lines indicate expected theoretical distribution. For TRF, $p(\theta)$ is a constant; for D05, $p(\theta)$ is biased towards $\theta = 0$ by a factor $\cos^N \theta$.

2.4.1.2 The Correlation Coefficient R^2_{xy} as a Measure of Invertibility

The subject of invertibility can also be approached via the well-known Pearson correlation coefficient R^2_{xy} (Pearson, 1896). R^2_{xy} can be interpreted as a measure of invertibility or non-invertibility of D05, or any other statistic under an inversion of the x- and y-axes (see, e.g., Rodgers & Nicewander, 1988). First, consider the case of fitting a line to a set of N data points $\{x_n, y_n\}$, with no intrinsic uncertainties (i.e., zero error bars) and no consideration of extrinsic scatter. In this case, linear regression gives the slope m_{xy} of the best-fit line to y vs. x:

$$m_{xy} = \frac{N\sum(x_n y_n) - \sum x_n \sum y_n}{N\sum(x_n^2) - (\sum x_n^2)}\,, \tag{2.33}$$

while the slope m_{yx} of the best-fit line to x vs. y (as measured in the inverted space) is:

$$m_{yx} = \frac{N\sum(x_n y_n) - \sum x_n \sum y_n}{N\sum(y_n^2) - (\sum y_n^2)}\,, \tag{2.34}$$

where the summations are all over the range $n = \{1 \ldots N\}$. In this simple linear regression case, Pearson's correlation coefficient R^2_{xy} can be expressed as:

$$R^2_{xy} \equiv \frac{\left[N\sum(x_n y_n) - \sum x_n \sum y_n\right]^2}{\left[N\sum(x_n^2) - (\sum x_n)^2\right]\left[N\sum(y_n^2) - (\sum y_n)^2\right]}\,. \tag{2.35}$$

By Equations 2.33, 2.34 & 2.35, Pearson's correlation coefficient is thus:

$$R^2_{xy} \equiv m_{xy} m_{yx}\,. \tag{2.36}$$

The above derivation does not take into account any intrinsic error bars or extrinsic slop in the data. However, we can use Equation 2.36 to generalize the R^2_{xy} measure, where m_{xy} and m_{yx} may come from linear fits with different statistics to data with both intrinsic and extrinsic scatter. Consider again the simple example of linear fits to

slop-dominated Gaussian random clouds of §2.4.1.1. Given the bias of D05 towards the x-axis, we expect $m_{xy} \to 0$ and $m_{yx} \to 0$. Hence, D05 will yield a generalized value of $R^2_{xy} \to 0$; this is what we might expect, given the clearly uncorrelated nature of these data sets.

More generally, for linear fits to any data set with the D05 statistic, $R^2_{xy} \leq 1$ always; this, in itself, is a consequence of the bias of D05 to shallower slopes discussed in the context of the Gaussian random clouds. If $R^2_{xy} \approx 1$, the results of the D05 fit can be trusted, even though the statistic is technically non-invertible. But the lower the value of R^2_{xy}, the less the fits of D05 (or any other non-invertible statistic) can be trusted.

For invertible statistics like R01 and TRF, since $m_{xy} = 1/m_{yx}$, $R^2_{xy} \equiv 1$ by definition. This means that the fitted results can always be trusted under inversion.[6] The fact that $R^2_{xy} \equiv 1$ does not mean that using TRF or R01 somehow magically makes a data set fully correlated. Rather, any lack of correlation will be reflected in the fitted uncertainty in m_{xy} (see §2.6.4), instead of in an R^2_{xy} value. It is important to note that this uncertainty in m_{xy} can be usefully propagated into other analyses, perhaps as part of a prior. A low value of R^2_{xy}, on the other hand, is useful only as a warning that the fitted result should not be trusted.

2.4.2 Behavior in 1D Limit

Another desirable property of any two-dimensional statistic is that it reduce to a χ^2-like statistic in the one-dimensional limits $\Sigma_{x,n} \to 0$ (all uncertainty in the y-direction) or $\Sigma_{y,n} \to 0$ (all uncertainty in the x-direction). In the former case, we require that:

$$\mathcal{L} \to \prod_{n=1}^{N} \frac{1}{\Sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{y,n}}{\Sigma_{y,n}}\right)^2\right], \tag{2.37}$$

[6]Note, however, that these statistics may have other problems the user should be aware of. For instance, unlike D05, TRF fits are not invariant under a multiplicative scaling of the x- or y-axis; we discuss this behavior §2.4.4, and introduce an alternative, scale-based correlation coefficient in §2.5.1.

and, in the latter, that:

$$\mathcal{L} \to \prod_{n=1}^{N} \frac{1}{\Sigma_{x,n}} \exp\left[-\frac{1}{2}\left(\frac{\delta_{x,n}}{\Sigma_{x,n}}\right)^2\right], \tag{2.38}$$

where $\delta_{y,n}$ and $\delta_{x,n}$ are the distances between the data point centroid (x_n, y_n) and the model curve in the y- and x-directions, respectively. Again, we refer to these statistics as "χ^2-like", because they are analogous to the traditional one-dimensional χ^2 statistic, with the caveat that Σ_n is not fixed, but a function of a free model parameter, i.e., the slop in the relevant dimension (§2.3). It should be obvious by inspection that the argument of the exponentials in the D05, R01 and TRF likelihood functions (Equations 2.20, 2.21 & 2.22) satisfy both Equation 2.37 and 2.38 in both limiting cases. If $\Sigma_{x,n} = 0$, the tangent point will lie along the curve $y_c(x; \vartheta_m)$ at the x-value of the data point, i.e., at $(x_n, y_c(x_n; \vartheta_m))$. And if $\Sigma_{y,n} = 0$, the tangent point will lie along the curve at the y-value of the data point, i.e., at $(y_c^{-1}(y_n; \vartheta_m), y_n)$. Again, the behavior of the statistics is entirely determined by the form of the prefactor in the likelihood function.

It should also be obvious by inspection that D05 (Equation 2.20) reduces to Equation 2.37 when $\Sigma_{x,n} \to 0$, but that it does not reduce to Equation 2.38 when $\Sigma_{x,n} \to 0$; in the latter case, there is an extra factor of $1/m_{t,n}$ in the prefactor. To use D05 to fit to data in this one-dimensional limit, one must first invert the x- and y-axes; in this case, the inverted D05* statistic of Equation 2.30 reduces to Equation 2.38. TRF has the advantage that one does not need to keep track of whether one-dimensional error bars are in the x- or y-direction and perform the inversion; it works, automatically (Equation 2.22 reduces to Equations 2.37 & 2.38 in both limiting cases).

Note, however, that R01 (Equation 2.21) does not reduce to the one dimensional statistics in either limit; there are extra factors of $\sqrt{1 + m_{t,n}^2}$ when $\Sigma_{x,n} \to 0$, and

$\sqrt{1 + 1/m_{t,n}^2}$ when $\Sigma_{y,n} \to 0$. We consider this a fatal flaw of R01; it was one of the motivating factors in developing a new invertible statistic in the first place.

2.4.3 Do Two Points Really Define a Line?

As a very basic reality check, we investigate the seemingly trivial problem of fitting a line to two data points. We proceeded under the assumption that, for a two-point data set, the concept of extrinsic scatter is meaningless, i.e., the slop parameters (σ_x, σ_y) are held fixed at zero. In this very simple case, our intuition leads us to expect the best-fit line to pass through the centroids of both data points' error ellipses; "two points define a line," after all. Or do they? We find that *none* of the three statistics, neither D05, R01, nor TRF, is capable of producing this intuitive result in all cases.

Figure 2.11 shows a series of linear fits to two data points with centroids (x_n, y_n) at $(-1, 1)$ and $(1, 1)$. The intrinsic error bars in each panel are a function of the parameter ρ, such that $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$; $\rho = 0°$ corresponds to the one-dimensional case $\sigma_{x,n} = 0$, $\sigma_{y,n} = 1$, while $\rho = 90°$ corresponds to the one-dimensional case $\sigma_{x,n} = 1$, $\sigma_{y,n} = 0$. The D05 statistic recovers the expected $\theta = 45°$ line only in the $\rho = 0°$ case, while the inverted D05* statistic succeeds only for $\rho = 90°$. The R01 statistic succeeds only in the $\rho = 45°$ case (i.e., when $\sigma_{x,n} = \sigma_{y,n}$). TRF, on the other hand, recovers $\theta = 45°$ for $\rho = 0°, 45°$ and $90°$. This result is not surprising, in light of the fact that TRF reduces to D05 (or inverted D05*) in both one-dimensional cases ($\Sigma_{x,n} \to 0$, $\Sigma_{y,n} \to 0$; § 2.4.2), and to R01 in the case where $\Sigma_{x,n} = \Sigma_{y,n}$. Figure 2.12 shows fitted θ vs. ρ for D05, inverted D05*, R01 and TRF. So, while none of these statistics is "perfect" in light of this two-point test, TRF at least succeeds, by itself, where all the others succeed separately.

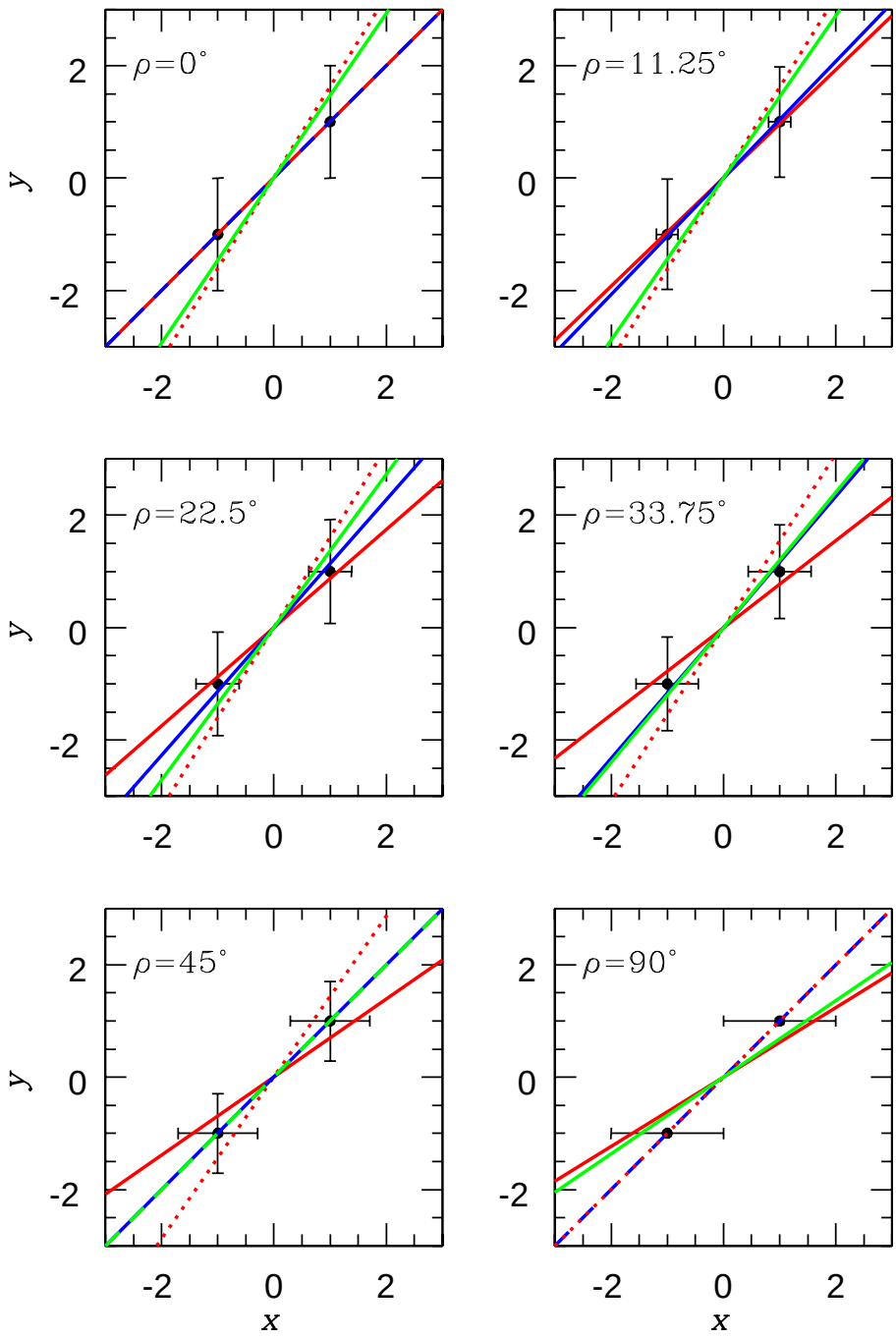

Figure 2.11 Linear fits to two data points, with centroids at $(-1, -1)$ and $(1, 1)$, with a range of intrinsic uncertainties, $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$. Solid red line: D05. Dotted red line: inverted D05*. Green line: R01. Blue line: TRF.

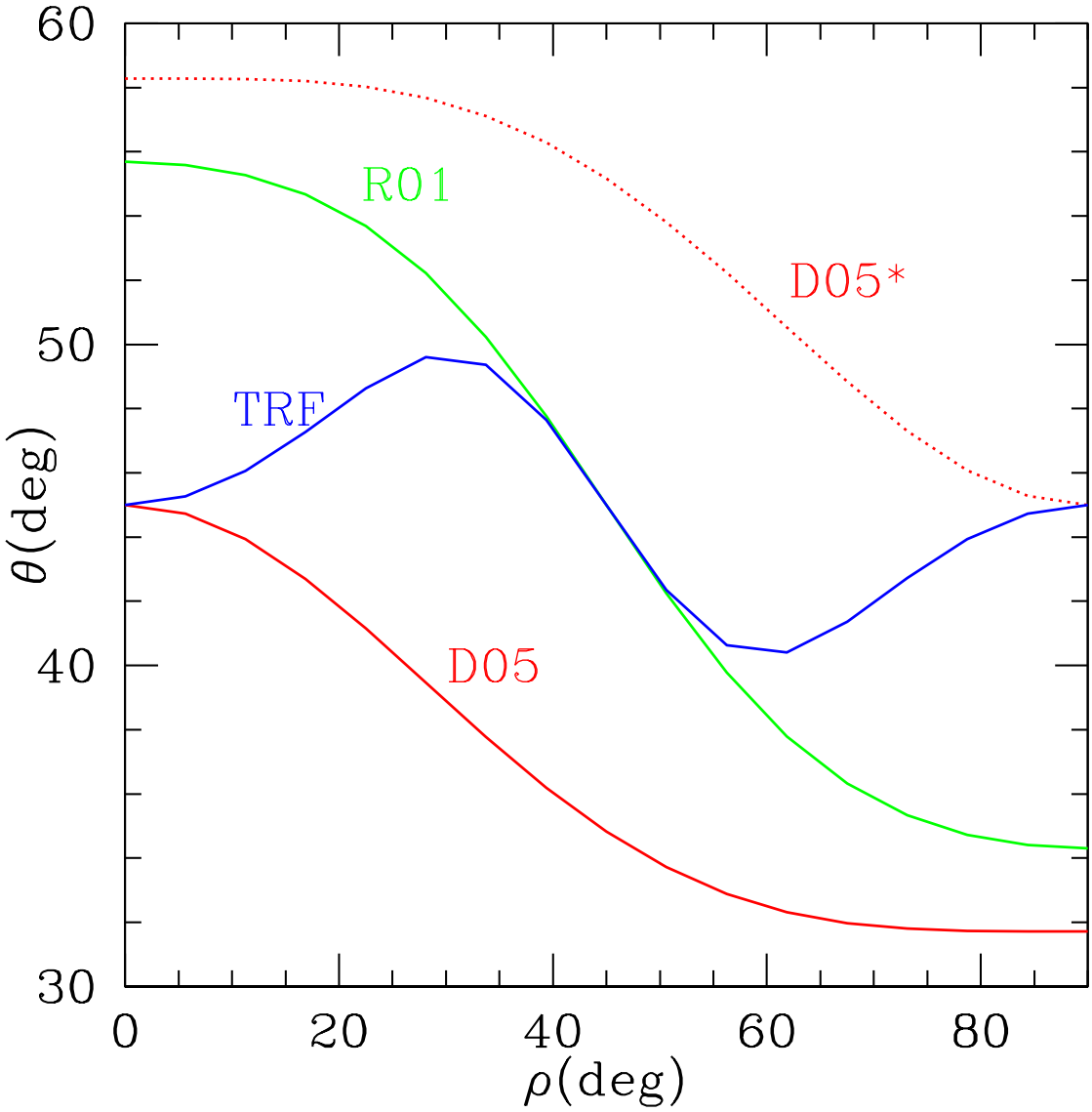

Figure 2.12 Best-fit linear position angle θ vs. axial ratio parameter ρ for two-point data sets with centroids at $(-1,-1)$ and $(1,1)$, for D05, inverted D05*, R01 and TRF statistics. Parameter ρ describes intrinsic uncertainties of each data point: $\sigma_{x,n} = \sin\rho$, $\sigma_{y,n} = \cos\rho$.

2.4.4 Scalability

While the TRF statistic is invariant under an inversion of the x- and y-axes, reduces to a χ^2-like statistic in the 1D limit, and passes the two-point test in more cases than either D05 or R01, it is not invariant under an arbitrary rescaling of the x- or y-axes; in other words, the best-fit that TRF provides depends on the choice of units of measurement (or on the choice of bases if fitting to data in logarithmic space). Since there is no a priori reason that one set of units (or bases) is preferable to another, it is therefore necessary to explore the behavior of TRF under all arbitrary, physically meaningful scalings in order to fully describe its predictions.

It is obvious, by inspection, that the R01, D05 and TRF statistics are invariant if both the x- and y-axes are rescaled by the same factor. So, without loss of generality, we can explore the full range of scale-dependent behavior of these statistics by considering only rescalings of the y-axis by a factor s (equivalently, we can define the scaling factor $s \equiv s_y/s_x$; the results are the same). The D05 statistic is, in fact, fully scalable. If we substitute $y_n \to sy_n$, $y_{t,n} \to sy_{t,n}$, $\Sigma_{y,n} \to s\Sigma_{y,n}$, and $m_{t,n} \to sm_{t,n}$ in Equation 2.20, we obtain the scaled likelihood function:

$$\begin{aligned}
\mathcal{L}_s^{\mathrm{D05}} &\propto \prod_{n=1}^{N} \frac{1}{s\sqrt{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2}} \exp\left\{ -\frac{1}{2} \frac{[y_n - y_{t,n} - m_{t,n}(x_n - x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2} \right\} \\
&= \left(\frac{1}{s^N}\right) \mathcal{L}^{\mathrm{D05}} \\
&\propto \mathcal{L}^{\mathrm{D05}} \quad \text{for a fixed } s\,. \qquad (2.39)
\end{aligned}$$

The argument of the exponential is invariant under the scaling; note, again, that this argument is the same for all three statistics we are considering here. The only effect of the scaling is the introduction of a constant factor $1/s$ in the prefactor. Therefore,

D05 will yield exactly the same best-fit curve (when plotted in the original, unscaled space); the total fitness $-2\ln\mathcal{L}_s^{\mathrm{D05}}$ will simply differ from $-2\ln\mathcal{L}^{\mathrm{D05}}$ by a constant offset of $2N\ln s$. In other words, $-2\ln\mathcal{L}_s^{\mathrm{D05}} - (-2\ln\mathcal{L}_s^{\mathrm{D05}})_{\min}$ will differ from $-2\ln\mathcal{L}^{\mathrm{D05}} - (-2\ln\mathcal{L}^{\mathrm{D05}})_{\min}$ by zero.

However, the TRF likelihood of Equation 2.22 scales as:

$$\begin{aligned}\mathcal{L}_s^{\mathrm{TRF}} &\propto \prod_{n=1}^{N}\sqrt{\frac{m_{t,n}^2\Sigma_{x,n}^2+\Sigma_{y,n}^2}{m_{t,n}^2\Sigma_{x,n}^4+s^2\Sigma_{y,n}^4}}\exp\left\{-\frac{1}{2}\frac{[y_n-y_{t,n}-m_{t,n}(x_n-x_{t,n})]^2}{m_{t,n}^2\Sigma_{x,n}^2+\Sigma_{y,n}^2}\right\}\\ &\not\propto \mathcal{L}^{\mathrm{TRF}} \quad \text{for a fixed } s\,. \end{aligned} \tag{2.40}$$

That is, the scaling factor s cannot be factored out of the likelihood function. Consequently, different choices of s will result in different fitted values of the slop parameters (σ_x, σ_y), and, through $m_{t,n}$, the other model parameters ϑ_m. (It can be shown, similarly, that $\mathcal{L}^{\mathrm{R01}}$ does not scale either, but we do not investigate the behavior of R01 any further in this discussion, given its other limitations.) Furthermore, it is impossible to use the numerical value of the likelihood function to make any assertions as to the relative fitness of the best fits that TRF yields at different scales. In other words, the scaling factor s is not a model parameter that can fit to. Its value cannot be determined by the data, but is rather a choice that must be made before fitting. The primary practical effect is that the best-fit model distribution, plotted in the original, unscaled space, is different for different choices of s.

Consider the scale dependence of slop direction, i.e., how the total slop is distributed between σ_x and σ_y. In the TRF statistic, this is most vividly illustrated in the case of linear fits to slop-dominated data, i.e., when $(\sigma_{x,n}, \sigma_{y,n}) = 0$. In the limit $s \to 0$, the likelihood of Equation 2.40 is dominated by a term of order $1/\sigma_x$ in the prefactor. To

maximize the likelihood, the TRF fit will force $\sigma_x \to 0$, and Equation 2.40 reduces to:

$$\mathcal{L}^{\mathrm{TRF}}_{s\to 0} \to \prod_{n=1}^{N} \frac{1}{s\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\delta_{y,n}}{\sigma_y}\right)^2\right]. \tag{2.41}$$

In other words, in the limit $s \to 0$, TRF reduces to the one-dimensional, χ^2-like statistic of Equation 2.37, modulo a constant factor of $(1/s)^N$. In the limit $s \to \infty$, the likelihood of Equation 2.40 is dominated by a term of order $1/s\sigma_y$ in the prefactor. To maximize the likelihood, the TRF fit will force $\sigma_y \to 0$, and Equation 2.40 reduces to:

$$\mathcal{L}^{\mathrm{TRF}}_{s\to \infty} \to \prod_{n=1}^{N} \frac{1}{\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{\delta_{x,n}}{\sigma_x}\right)^2\right]. \tag{2.42}$$

In other words, for linear fits to slop-dominated data, in the limit $s \to \infty$, TRF reduces to the one-dimensional, χ^2-like statistic of Equation 2.38.

Another way of looking at this scaling behavior is that, for linear fits to slop-dominated data, *the predictions of TRF at these two extremes of scale are equivalent to those of D05 under inversion.* As $s \to 0$, TRF yields the same results as D05, and as $s \to \infty$, it yields the same results as inverted D05*. However, there are two significant advantages to using TRF: (1) TRF can explore a *continuum* of scales in between those limits, while D05 is restricted to the binary choice of inversion or non-inversion. In §2.5, we will introduce a method for selecting an optimum midrange scale for TRF fits. (2) If the intrinsic error bars $(\sigma_{x,n}, \sigma_{y,n})$ are non-zero, the *physically meaningful range* of the scale continuum is a subset of $[0, \infty]$. As the intrinsic scatter grows relative to the extrinsic scatter, this range of physically meaningful scales shrinks. In the limit where the data are entirely dominated by intrinsic scatter (i.e., both slops equal zero), the range reduces to a single physically meaningful scale, with a single best-fit solution.

2.4.5 Degeneracy of Slop for Linear Models

In fitting linear models to data with extrinsic scatter, intuition tells us that there should be a degeneracy between slop in the x- and y-directions. That is, if the fitted line has slope m, the model distribution should be equally well-described by putting slop σ entirely in the x-direction, or by putting slop $m\sigma$ entirely in the y-direction, or into any combination of the two such that $\sigma^2 \equiv m^2\sigma_x^2 + \sigma_y^2$. All three statistics (Equations 2.20, 2.21, & 2.22) have the same argument in the exponential function, with a term in the denominator:

$$\begin{aligned} \Sigma_n^2 &\equiv m^2\Sigma_{x,n}^2 + \Sigma_{y,n}^2 \\ &\equiv m^2\sigma_{x,n}^2 + \sigma_{y,n}^2 + m^2\sigma_x^2 + \sigma_y^2 \\ &\equiv m^2\sigma_{x,n}^2 + \sigma_{y,n}^2 + \sigma^2 \,. \end{aligned} \tag{2.43}$$

That is, the quantity Σ_n, and therefore the argument of the exponential for all three statistics, is degenerate in slop for linear fits.

The D05 statistic (Equation 2.20) has a prefactor that is simply $1/\Sigma_n$, and so linear fits with D05 are fully degenerate in slop: it cannot be determined in which direction the slop lies. Similarly, the R01 (Equation 2.21) prefactor is $\sqrt{1+m^2}/\Sigma_n$, which also preserves slop degeneracy. The prefactor of TRF (Equation 2.22), on the other hand, is equal to $\Sigma_n/\sqrt{m^2\Sigma_{x,n}^4 + \Sigma_{y,n}^4}$. In this case, the degeneracy is broken, meaning that TRF will pick specific values for σ_x and σ_y, even in the slop-dominated linear case. We saw this in §2.4.4: In the limit $s \to \infty$, all slop is placed in the x-direction ($\sigma_y \to 0$), and in the limit $s \to 0$, all slop is placed in the y-direction ($\sigma_x \to 0$). This does *not* mean that TRF can magically disentangle σ_x from σ_y. Rather, how much slop is put into one direction or the other depends on the choice of scale. The non-scalability of TRF and its non-degeneracy of slop in linear fits are, in a very real sense, equivalent.

Since we can conduct a particular fit only at a fixed scale, it is important to keep in mind that, when using TRF, we should not interpret the best-fit values of (σ_x, σ_y) *physically*. Other values of slop in either direction are possible, given other choices of scale. We must instead view the fitted slops as an *empirical* parameterization of extrinsic scatter in the data at that particular scale. This is not a problem, considering that this is how we introduced slop in the first place (§2.2): as an empirical parameterization of extrinsic scatter about the curve $y_c(x; \vartheta_m)$ that has a physical explanation, albeit one that we cannot or choose not to model physically. For linear fits, the model distribution obtained by convolving the curve $y_c(x; \vartheta_m)$ with the two-dimensional Gaussian that represents the slop is the same, whether the slop is entirely in the x-direction, entirely in the y-direction, or in some combination of the two. There is no consequence of the loss of slop degeneracy, except perhaps that it differs from our intuitive experience with χ^2-like statistics. It simply reflects the fact that, in TRF fits, the directionality of slop, as well as its magnitude, is empirical. For non-linear fits, in cases where the data require that there be some slop in both directions, the best-fit TRF curves at extremes of scale may jump to unphysical regimes in order to accommodate all the data with only one non-zero slop parameter, and so the physically meaningful range of scales may be reduced (see, e.g., Figure 2.19 in §2.5).

2.4.6 Summary of Strengths and Weaknesses

Table 2.1 summarizes the properties of the D05, R01 and TRF statistics discussed above. Clearly, each statistic has its own strengths and weaknesses. We consider TRF to be, on the whole, an improvement on R01; R01 is, in fact, just a special case of TRF when $\Sigma_{x,n} = \Sigma_{y,n}$, and will not be discussed further here. TRF has the advantage of invertibility over D05 (and reduces to D05 in both one-dimensional limits). On the other hand, D05 is invariant under scaling, while TRF is not. Also, in the case of

Table 2.1. Summary of Statistical Properties

	D05	R01	TRF
Invertible?	No	Yes	Yes
Reduces to 1D χ^2?	Yes	No	Yes
Passes 2-point test?	If $\Sigma_x = 0$	If $\Sigma_x = \Sigma_y$	If $\Sigma_x = 0$, $\Sigma_y = 0$, or $\Sigma_x = \Sigma_y$
Scalable?	Yes	No	No
Slop degenerate if linear?	Yes	Yes	No

TRF, slops fitted at a single scale cannot be interpreted physically (§2.4.5), but only as empirical parameterizations of extrinsic scatter in the data at that scale. However, the advantage of TRF is that we can explore cases of intermediate scale, while D05 is restricted to a binary choice of inversion or non-inversion. And, as we will discuss further in §2.5, the range of fits of TRF under extremes of scale are identical to, or a *subset* of, those of D05 under inversion.

2.5 Implementation of the TRF Statistic

2.5.1 A New Correlation Coefficient for the TRF Statistic

As we discuss in §2.4.4, for slop-dominated linear data (i.e., intrinsic error bars equal zero), the behavior of TRF under extremes of scale is equivalent to the behavior of D05 under an inversion of the x- and y-axes. Specifically, as the scaling factor $s \to 0$, TRF becomes equivalent to the one-dimensional D05 statistic of Equation 2.37, and as $s \to \infty$, it is equivalent to the one-dimensional inverted D05* statistic of Equation 2.38. Figure 2.13 illustrates this limiting behavior. Data points were generated by adding independent zero-mean Gaussian random variables with standard deviation $\sigma_x = \sigma_y = 0.5$ to points randomly distributed along the line $y_c(x) = x$. The dashed-dotted lines

show the non-inverted D05 fit (shallower line), and inverted D05* fit (steeper line). These fits are identical to the TRF best-fits in the scale limits $s \to 0$ and $s \to \infty$, respectively. With the D05 statistic, these two lines are the only possible fits to this data set. However, with TRF, we have the freedom to choose any intermediate scale s along a continuum. The solid blue line shows the TRF fit at an "optimum" intermediate scale, which we will discuss further below.

Figure 2.14 shows fits to the identical data set as in Figure 2.13, but with intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0.25$ assigned to each point. The dotted red lines show the D05 and inverted D05* fits, which are *identical* to those for the slop-dominated data set Figure 2.13. The dashed blue lines show TRF fits at the scales $s_{\min}$ and $s_{\max}$ at which the fitted slops $\sigma_x \to 0$ and $\sigma_y \to 0$, respectively. Clearly, in this case, TRF predicts a narrower range of fitted slopes than does D05.

For fits to data with non-zero intrinsic uncertainties with the TRF statistic, fits that fall outside of the range obtained at these two limiting scales would require either that $\sigma_x^2 < 0$ (at smaller scales) or $\sigma_y^2 < 0$ (at larger scales). Consider the extreme limiting cases of $s \to 0$ and $s \to \infty$ in the linear fits shown in Figure 2.14. In the former case, in order to reproduce the fit obtained with D05, TRF would have to reduce to the one-dimensional statistic of Equation 2.37. The only way this could happen would be if the quantity $\Sigma_n^2 \equiv m^2(\sigma_{x,n}^2 + \sigma_x^2) + (\sigma_{y,n}^2 + \sigma_y^2)$ (Equation 2.43) were to reduce to $\sigma_{y,n}^2 + \sigma_y^2$, which would require the slop parameter $\sigma_x^2 < 0$ (for correlated data with non-zero slope). Similarly, in the latter case, to reproduce the inverted D05* fit, TRF would have to reduce to the one-dimensional statistic of Equation 2.38, which would require the slop parameter $\sigma_y^2 < 0$. We defined the model distribution as the convolution of a two-dimensional Gaussian with a model curve, where the 1σ widths of the Gaussian in the x- and y-directions are equal to (σ_x, σ_y). Simply put, there is no real, physically meaningful model distribution in the x-y plane that can be constructed with imaginary

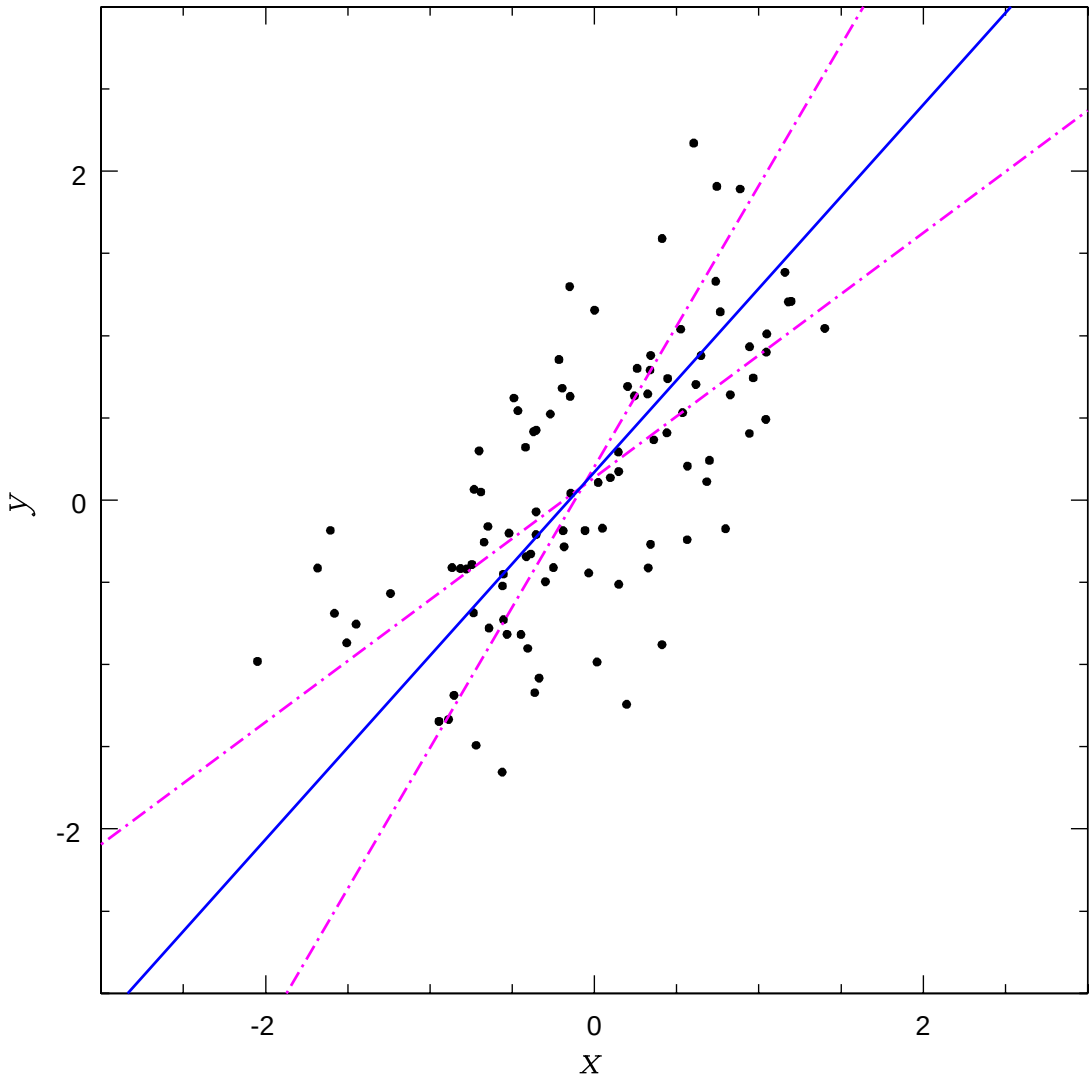

Figure 2.13 Linear fits to slop-dominated data. Data points generated by adding independent zero-mean Gaussian random variables with standard deviation $\sigma_x = \sigma_y = 0.5$ to $N = 100$ points randomly distributed between $(-1, -1)$ and $(1, 1)$ along the line $y_c(x) = x$. Data points are assigned zero intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0$. Dash-dotted lines: uninverted D05 fit/TRF fit as scale $s \to 0$ (shallow line), and inverted D05* fit/TRF fit as scale $s \to \infty$ (steeper line). Solid blue line: TRF fit at optimum scale.

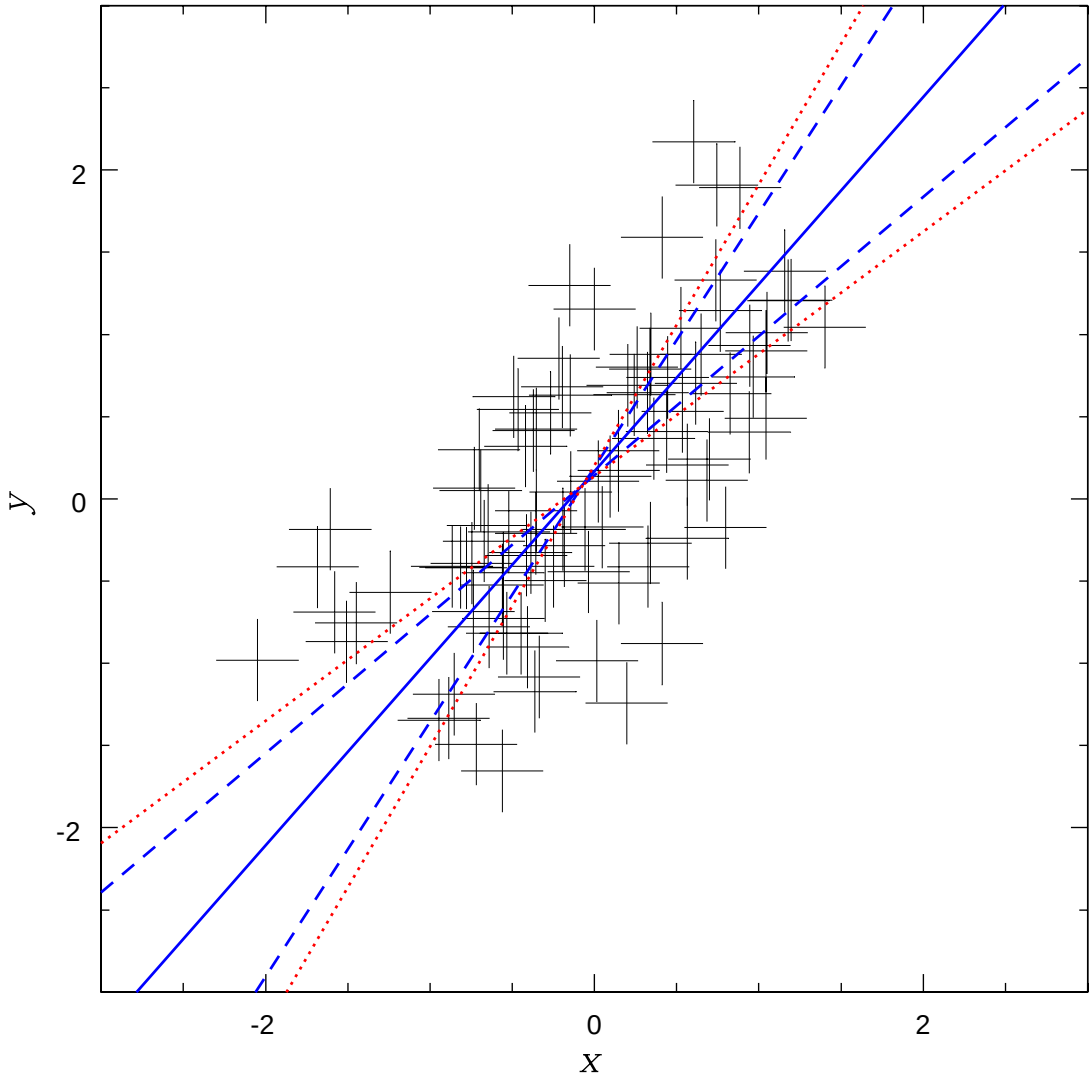

Figure 2.14 Linear fits to data with both slop and intrinsic uncertainties. The coordinates of each point are the same as in Figure 2.13. Each point was assigned symmetric intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0.25$. Dotted red lines: uninverted D05 fit (shallow line), and inverted D05* fit (steeper line); these fits are identical to those in Figure 2.13. Dashed blue lines: TRF fits at scale $s_{\min}$ at which fitted slop $\sigma_x \to 0$ (shallower line), and at scale $s_{\max}$ at which slop $\sigma_y \to 0$ (steeper line). Solid blue line: TRF fit at optimum scale.

values of slop. Therefore, any scenario that requires $\sigma_x^2 < 0$ or $\sigma_y^2 < 0$ is unphysical, and the range $[s_{\min}, s_{\max}]$ represents the full, physically meaningful range of scales for fits with the TRF statistic.[7]

In general, in fits to data with intrinsic error bars, TRF will predict a narrower range of best-fit curves under physically meaningful changes of scale than D05 will under inversion. How do we quantify this difference? As we discussed in §2.4.1.2, the linear correlation coefficient R_{xy}^2 can be interpreted as a measure of the reliability of a statistic under an inversion of the x- and y-axes. Specifically, for a linear fit, if the slope of the best-fit line to y vs. x is m_{xy}, and the slope of the line fit to x vs. y is m_{yx} (as measured in the inverted space), then we can define a generalized correlation coefficient $R_{xy}^2 \equiv m_{xy} m_{yx}$. From this definition of R_{xy}^2 as the product of the slopes of best-fit lines under an inversion of the x- and y-axes, it should be clear that for an invertible statistic like TRF, $R_{xy}^2 \equiv 1$ for all linear fits, since $m_{xy} = 1/m_{yx}$, by definition of invertibility. Thus, for an invertible statistic, R_{xy}^2 can provide no statistical information.

We seek an analogous "correlation coefficient" for TRF, one that is a measure of the variance of the statistic's predictions under a change of scale. We require that this coefficient share the properties of the traditional R_{xy}^2, i.e., that $0 \leq R_{\rm TRF}^2 \leq 1$, where $R_{\rm TRF}^2 = 1$ corresponds to best-fit lines of identical slope (as plotted in the same scale space) for the two scales we are comparing, while $R_{\rm TRF}^2 = 0$ corresponds to best-fit lines that are orthogonal. One option, if we were only concerned with linear fits, would be to define an analogous correlation coefficient in terms of the ratio of the slopes of the best-fit lines at the two extremes of scale. But even in the linear case, there are

[7]There remains the possibility that if, at a given scale, a fit requires $\sigma_x^2 < 0$ or $\sigma_y^2 < 0$, then the intrinsic error bars $\sigma_{x,n}$ or $\sigma_{y,n}$ are overestimated. If there is genuine cause for concern that the quoted statistical error bars for a given data set are overestimated, we technically have the option to introduce, as free model parameters, new multiplicative scaling factors $(f_x, f_y) \leq 1$ to the intrinsic error bars whenever $(\sigma_x^2, \sigma_y^2) < 0$, i.e., to replace $(\sigma_{x,n}, \sigma_{y,n})$ with $(f_x \sigma_{x,n}, f_y \sigma_{y,n})$. In this case, TRF will asymptote to the limiting behavior of one-dimensional D05, or inverted D05*, as $f_x \to 0$, or $f_y \to 0$, respectively. We do not explore this option further here, i.e., we assume that the given intrinsic error bars in the data sets we analyze are not overestimated, at least not significantly.

scenarios where this definition would fail. For instance, if the best-fit TRF line at one scale extreme has a slope $m = 0$, and at the other extreme has a slope just slightly greater or less than zero, any correlation coefficient defined in terms of the ratios of the slopes would either equal zero or $\pm\infty$ – despite the fact that the two limiting lines are quite similar.

Note that this divergence of R^2_{xy} never occurs with fits using the D05 statistic. For instance, consider a linear fit to uncorrelated data, i.e., a fit for which D05 gives a slope $m \approx 0$. For such a data set, whether the data are dominated by extrinsic scatter or intrinsic uncertainty, the inverted D05* fit will give a slope $m \to \infty$, and hence a correlation coefficient $R^2_{xy} \approx 0$. If the data are correlated, but dominated by extrinsic scatter, likewise, TRF will reproduce the D05 and inverted D05* fits at the extremes of scale $s \to 0$ and $s \to \infty$, respectively, as discussed above. However, if the data are dominated by intrinsic uncertainty, TRF will return similar slopes over the full range of physically meaningful scales; in the special case of $m \approx 0$, R^2_{xy} may then diverge. Pearson's correlation coefficient has stood the test of time in no small part thanks to the fact that linear fits with non-invertible statistics never encounter this failing case. Fortunately, correlation coefficients can be defined in many ways: there is an abundance of examples in the literature.

We introduce a new correlation coefficient, R^2_{TRF} that is a function of the *difference* of the slopes, as measured by position angles, of the two curves we are comparing, rather than the slope *ratio*. First consider the case of a linear fit. Let the slope of the best-fit line at scale $s_{\min} = a$ be $m_a = \tan\theta_a$, and the slope of the best-fit line at scale $s_{\max} = b$ be $m_b = \tan\theta_b$, as plotted in the original scale space $s = 1$. Then we can define a scale-dependent correlation coefficient in terms of the difference $|\theta_a - \theta_b|$:

$$R'^2_{\mathrm{TRF}}(a,b) \equiv \tan^2\left(\frac{\pi}{4} - \frac{|\theta_a - \theta_b|}{2}\right). \tag{2.44}$$

By inspection, it is clear that $R'^2_{\rm TRF}(a,b) = 1$ if $\theta_a = \theta_b$, while $R'^2_{\rm TRF}(a,b) = 0$ if $\theta_a = \theta_b \pm \pi/2$, in the scale space $s = 1$. In order to find the "optimum scale" s_0, we consider the limiting behavior of the fits at the scales (a,b), defined either as $s \to 0$ and $s \to \infty$ (for slop-dominated data), or else as the scales at which TRF becomes unphysical. These scales are typically those for which TRF would require fitted slops $(\sigma_x^2, \sigma_y^2) < 0$, though there are cases where best fits beyond a particular scale may be deemed unphysical for other reasons (see Figure 2.19, below). We then follow an iterative approach. The first approximation of s_0 is the scale at which:

$$R'^2_{\rm TRF}(a, s_0) = R'^2_{\rm TRF}(s_0, b) \equiv R^2_{\rm TRF}\,, \tag{2.45}$$

where, again, the position angles θ describe the slopes in the original, unscaled space $s = 1$. We then shift from the $s = 1$ space to this $s = s_0$ space, where the position angles of the lines are transformed as $\theta \to \tan^{-1}(s_0 \tan\theta)$. We repeat the analysis in this space, finding a new scale s_0' at which $R'^2_{\rm TRF}(a, s_0') = R'^2_{\rm TRF}(s_0', b) \equiv R^2_{\rm TRF}$, i.e.:

$$\begin{aligned} R^2_{\rm TRF} &\equiv \tan^2\left(\frac{\pi}{4} - \frac{\left|\tan^{-1}(s_0\tan\theta_a) - \tan^{-1}(s_0\tan\theta_{s_0'})\right|}{2}\right) \\ &= \tan^2\left(\frac{\pi}{4} - \frac{\left|\tan^{-1}(s_0\tan\theta_{s_0'}) - \tan^{-1}(s_0\tan\theta_b)\right|}{2}\right)\,, \end{aligned} \tag{2.46}$$

where $\theta_{s_0'}$ is the position angle of the best-fit line at scale s_0', as measured in $s = 1$ space. We then set $s_0 = s_0'$, and repeat the procedure, scaling all fitted position angles into that s_0 space and finding a new value of s_0'. We find that s_0 and $R^2_{\rm TRF}$ converge to stable values to very high precision after only a few iterations. This final value of s_0 we take to be the *optimum scale* for TRF fits to the data set; it is at this optimum scale that we compute uncertainties in the fitted model parameters (see §§2.6.4 & 2.6.5).

2.5.2 Properties and Behavior of $R'^2_{\mathrm{TRF}}(a,b)$ and R^2_{TRF}

In the procedure described above, we defined both $R'^2_{\mathrm{TRF}}(a,b)$ (Equation 2.44), which is a measure of the difference between the TRF fits at the two physically meaningful extremes of scale, and R^2_{TRF} (Equation 2.45), which is a measure of the difference between the fit at the optimum scale s_0 and the fit at either extreme. We find that when $R'^2_{\mathrm{TRF}}(a,b)$ is defined in this way with this procedure, it has the additional, very fortuitous property that $R'^2_{\mathrm{TRF}}(a,b) = R^2_{xy}$ for slop-dominated linear data, when computed at the optimum scale s_0. To demonstrate this fact empirically, we provide Figures 2.15 & 2.16. These two figures compare linear fits to slop-dominated data with a variety of slopes, for both loosely correlated ($\sigma_x = \sigma_y = 0.5$) and tightly correlated ($\sigma_x = \sigma_y = 0.2$) data sets. In both figures, the original data set (with different scalings of the same zero-mean Gaussian random variables) used in Figure 2.13 is rotated through a variety of position angles, with the rotations calibrated such that the best-fit D05/unscaled TRF line in the first panel of each figure gives $\theta = 0$. Note that, while $R'^2_{\mathrm{TRF}}(a,b) = R^2_{xy}$ in all cases, $R^2_{\mathrm{TRF}} > R^2_{xy}$. This is simply a reflection of the fact that, with TRF, we are free to explore a continuum of scales between the two extremes, and to choose an optimum scale that gives a best-fit line that lies in between the limiting fits of D05/D05*.

Figure 2.17 shows the behavior of $R'^2_{\mathrm{TRF}}(a,b)$ and R^2_{TRF} for the same $m \approx 1$ data set as in Figures 2.13 & 2.14, for a range of intrinsic error bars. In the slop-dominated case (Figure 2.17a), as we discussed above, the limiting behavior of TRF as $s \to \infty$ and $s \to 0$ is the same as that of D05 under inversion; in this case, $R'^2_{\mathrm{TRF}}(a,b) = R^2_{xy} = 0.435$, where $R'^2_{\mathrm{TRF}}(a,b)$ is computed at the optimum scale for this data set ($s_0 = 0.8861$). In the intermediate cases (Figure 2.17b,c) the range of physically meaningful scales for TRF narrows as the intrinsic error bars become larger relative to the slop, and $R'^2_{\mathrm{TRF}}(a,b)$ and R^2_{TRF} increase accordingly. Figure 2.17d shows the limiting case of data

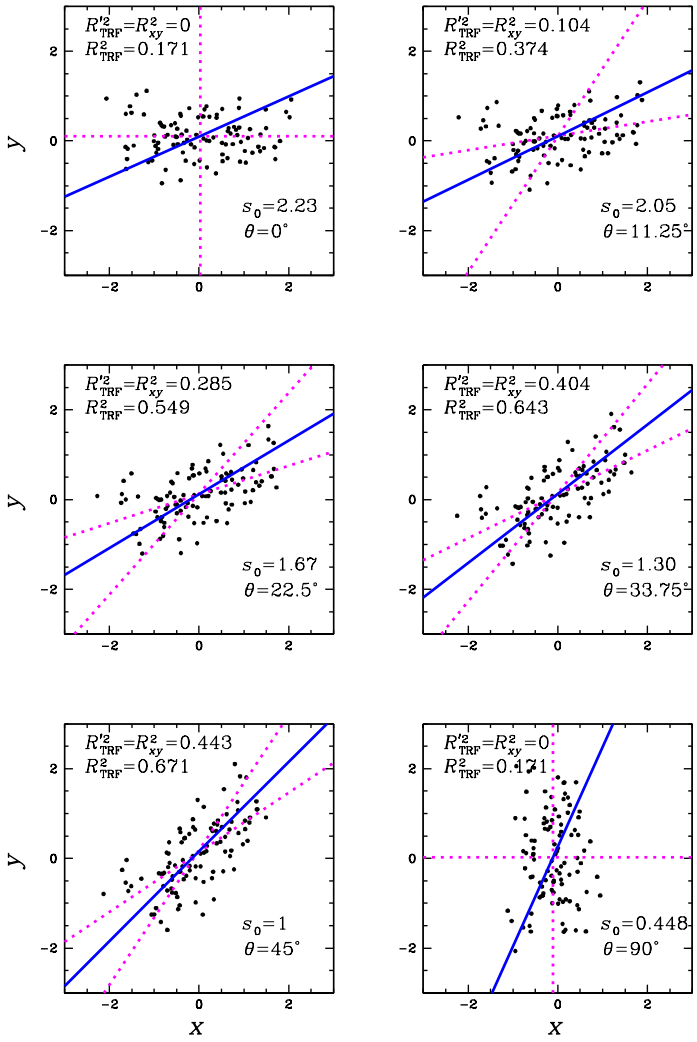

Figure 2.15 Equivalence of correlation coefficients $R'^2_{\rm TRF}(a,b)$ and R^2_{xy} for linear fits to loosely correlated ($\sigma_x \approx \sigma_y \approx 0.5$) slop-dominated data with a range of slopes. The coordinates of each point are generated by rotation of the data points in Figure 2.13 over a range of angles. Rotation angles were chosen so that the best-fit D05/unscaled TRF fit gives $\theta = 0$ in Figure (a). Dashed-dotted lines: uninverted D05/low-scale TRF fit (shallow line), and inverted D05*/high-scale TRF (steeper line). Solid blue lines: TRF fits at the optimum scale s_0. For slop-dominated data, $R'^2_{\rm TRF}(a,b) = R^2_{xy}$ at the optimum scale.

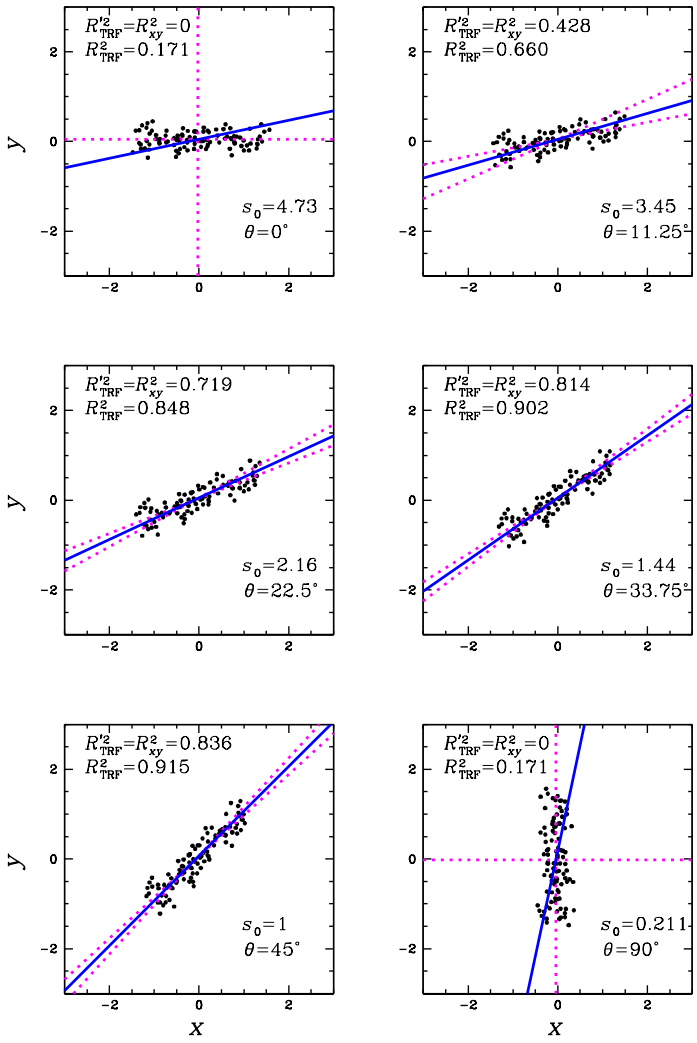

Figure 2.16 Equivalence of correlation coefficients $R'^2_{\rm TRF}(a,b)$ and R^2_{xy} for linear fits to tightly correlated ($\sigma_x \approx \sigma_y \approx 0.2$) slop-dominated data with a range of slopes. The coordinates of each point are generated from the same zero-mean Gaussian random variables and random $y_c(x) = x$ data points as in Figure 2.13, but with standard deviation $\sigma_x = \sigma_y = 0.2$, rotated over a range of angles. Rotation angles were chosen so that the best-fit D05/unscaled TRF fit gives $\theta = 0$ in Figure (a). Dashed-dotted lines: uninverted D05/low-scale TRF fit (shallow line), and inverted D05*/high-scale TRF (steeper line). Solid blue lines: TRF fits at the optimum scale s_0. For slop-dominated data, $R'^2_{\rm TRF}(a,b) = R^2_{xy}$ at the optimum scale.

entirely dominated by intrinsic error bars, i.e., TRF returns fitted slops $\sigma_x = \sigma_y = 0$ at scale $s_0 = 1$, and would require unphysical $(\sigma_x^2, \sigma_y^2) < 0$ for fits at all other scales, so that $R^2_{\rm TRF} \equiv 1$. In other words, for a data set dominated entirely by intrinsic uncertainty, the TRF statistic can be trusted completely; any uncertainty in the fitted slope will be fully reflected in the computed probability distributions of the slope parameter m (or position angle parameter $\theta = \tan^{-1} m$; see §2.6.4). Note that the behavior of D05 and inverted D05* is identical for all four cases, as is the value of $R^2_{xy} = 0.4348$: a value indicating that fits to these data sets with the D05 statistic should not be trusted. Finally, note that in all four cases, TRF fits at the optimum scale come closer to recovering the original $m = 1$ line from which these data sets were generated than either D05 or inverted D05*.

To summarize, for linear fits to entirely slop-dominated data, $R'^2_{\rm TRF}(a, b) = R^2_{xy}$ when computed at the optimum scale s_0; the calibration of our intuition about R^2_{xy} carries over to $R'^2_{\rm TRF}(a, b)$. The correlation coefficient $R^2_{\rm TRF}$ is always greater than $R'^2_{\rm TRF}(a, b)$, because it takes advantage of the fact that scale is a continuum, while invertibility is a bimodal choice. Furthermore, fits to entirely slop-dominated data sets represent TRF's worst-case scenario. For data with non-zero intrinsic error bars, TRF requires unphysical $(\sigma_x^2, \sigma_y^2) < 0$ in order to reproduce the fits of D05, and the range of physically meaningful scales is reduced. Increasing the intrinsic uncertainty with respect to the extrinsic scatter only increases the reliability of TRF; fits to data that are entirely dominated by intrinsic uncertainty can be trusted completely; any uncertainty in the fit will entirely be reflected in the computed probability distributions of the model parameters.

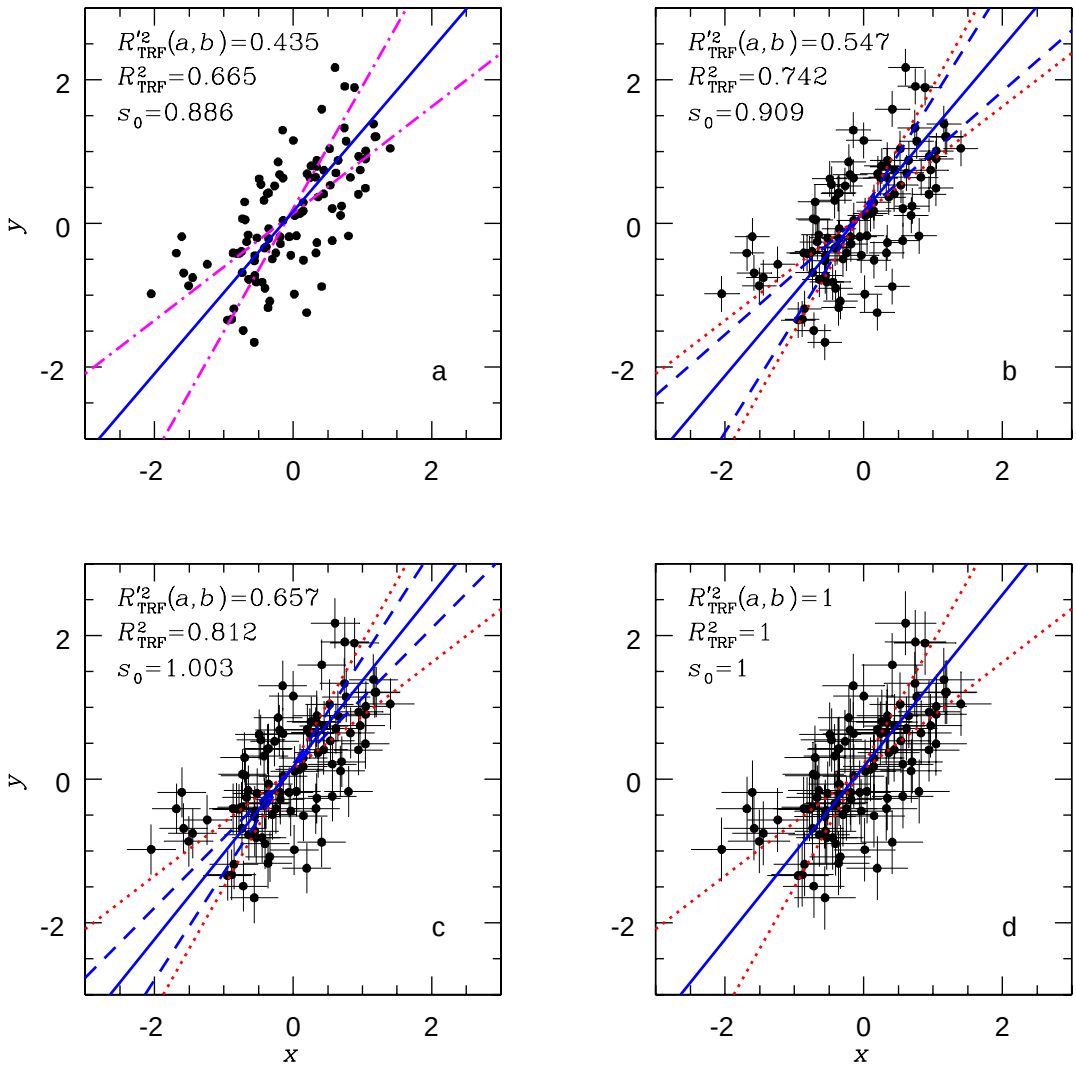

Figure 2.17 Comparison of correlation coefficients $R'^2_{\rm TRF}(a,b)$ and $R^2_{\rm TRF}$ for linear fits to $m \approx 1$ data with a range of intrinsic uncertainties. The coordinates of each point are the same as in Figure 2.13. Each point was assigned symmetric intrinsic uncertainties $\sigma_{x,n} = \sigma_{y,n} = 0, 0.25, 0.35, 0.4435$. Dotted red lines: uninverted D05 fit (shallow line), and inverted D05* fit (steeper line); these fits are identical in all cases to those in Figure 2.13. Dotted blue lines: TRF fits at scale $s_{\rm min}$ at which fitted slop $\sigma_x \to 0$ (shallower line), and at scale $s_{\rm max}$ at which slop $\sigma_y \to 0$ (steeper line). Solid blue line: TRF fit at optimum scale. (a) is the same as the slop-dominated case of Figure 2.13; in this case $R'^2_{\rm TRF}(a,b) = R^2_{xy} = 0.435$. (b) is the same as in Figure 2.14. (d) is the intrinsic uncertainty-dominated case, for which fitted slops in both directions, at all scales, equals zero, and $R^2_{\rm TRF} = 1$, meaning that the fit can be trusted completely; uncertainty in the slope will be fully reflected in its fitted probability distribution. Note that $R^2_{xy} = 0.435$ in all four cases.

2.5.3 TRF Fits to Non-Linear Data

For non-linear fits to a set of N data points, we generalize the linear form of $R'^2_{\mathrm{TRF}}(a,b)$ (Equation 2.44) as the average of N terms:

$$R'^2_{\mathrm{TRF}}(a,b) \equiv \frac{1}{N}\sum_{n=1}^{N} \tan^2\left(\frac{\pi}{4} - \frac{|\theta_{t,n;a} - \theta_{t,n;b}|}{2}\right), \tag{2.47}$$

Where $\theta_{t,n;a} = \tan^{-1} m_{t,n;a}$ and $\theta_{t,n;b} = \tan^{-1} m_{t,n;b}$ are the position angles of the best-fit curves at the tangent point to data point n at scales a and b, respectively. As with linear fits, we then iterate to find the optimum scale s_0 for which $R'^2_{\mathrm{TRF}}(a,s_0) = R'^2_{\mathrm{TRF}}(s_0,b) \equiv R^2_{\mathrm{TRF}}$.

Figure 2.18 illustrates the results of application of this iterative, pointwise, scale-dependent correlation coefficient to describe the non-linear correlation between observed interstellar extinction parameters R_V vs. c_2 for a sample of 441 stars in the Milky Way and Small and Large Magellanic Clouds, described in detail in Chapter 3. The solid line is the fit at the optimum scale s_0 for which $R'^2_{\mathrm{TRF}}(s_{\min},s_0) = R'^2_{\mathrm{TRF}}(s_0,s_{\max})$, as measured in the s_0 scaled space. The dotted line is the best TRF fit at the scale $s_{\min}$ at which slop $\sigma_x \to 0$. The dashed line is the best TRF fit at the largest scale $s_{\max}$ for which TRF recovers a physically meaningful curve.

This $s_{\max}$ is a smaller scale than the one at which the slop $\sigma_y \to 0$. Beyond $s_{\max}$, all best-fit curves, including local minima, discontinuously jump to a new family of curves where most of the data are fit by a line, but the right- or leftmost data are fit by a vertical segment (Figure 2.19). Such curves are the only way to accommodate the extrinsic scatter in the data with a model distribution that places slop primarily in the x-direction. Admittedly, since slop is purely an empirical parameterization of extrinsic scatter in the data at a given scale, these model *distributions* are empirically valid fits to the data at large scales. The problem is simply that the underlying *curves* are

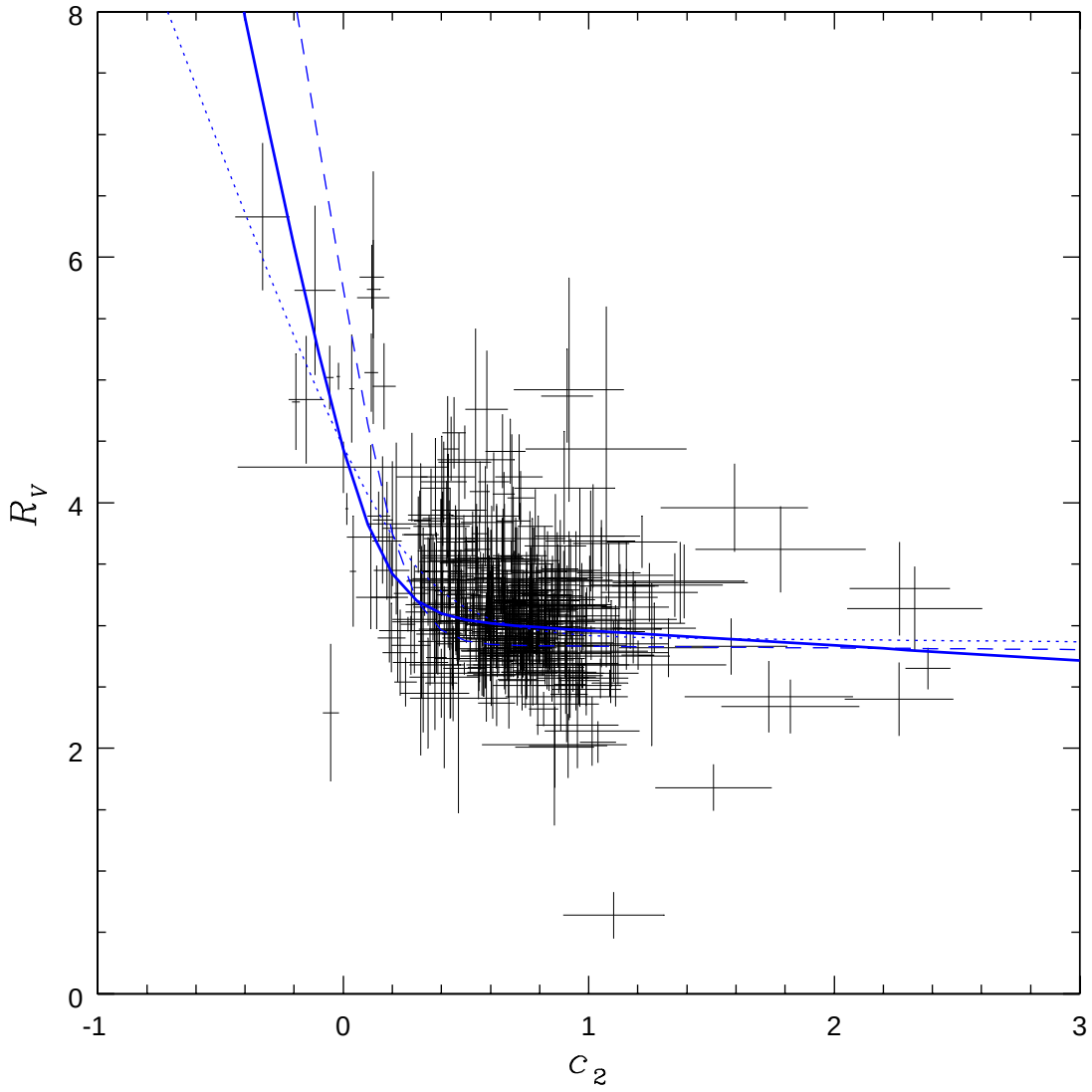

Figure 2.18 Example of TRF scale-dependence for a non-linear fit: observed interstellar extinction parameters R_V vs. c_2 (see §3.3). Dotted line: fit at $s_{\min}$, at which slop $\sigma_x \to 0$; Dashed line: fit at $s_{\max}$, the largest scale at which a physically meaningful best-fit curve can be obtained. Solid line is the fit at the scale s_0 for which $R'^2_{\rm TRF}(s_{\min}, s_0) = R'^2_{\rm TRF}(s_0, s_{\max})$. The best-fit slop envelopes for this model distribution are not shown here; see §3.3 for a more thorough discussion of this fit.

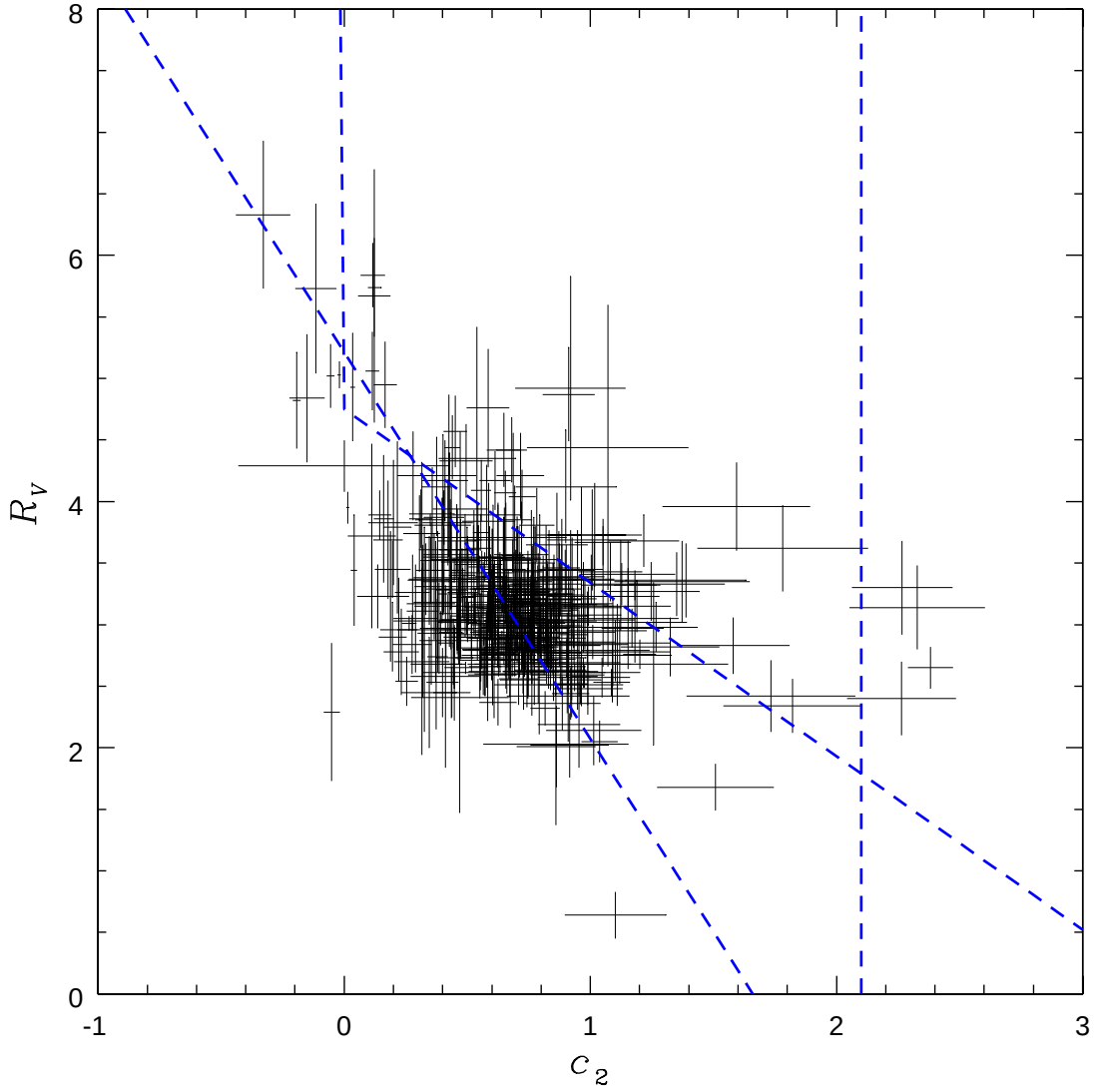

Figure 2.19 Examples of unphysical TRF fits to R_V vs. c_2 at scales $s > s_{\max}$, for which the slop $\sigma_y \to 0$. No physically plausible distributions of dust could result in such relations. The best-fit slop envelopes for this model distribution are not shown here; see §3.3 for a more thorough discussion of this fit.

unphysical, i.e., there is no plausible physical distribution of dust that could produce such relationships between R_V and c_2.

Finally, note that in the region $c_2 \gtrsim 1$, the slopes of the limiting curves in Figure 2.18 are close to zero. If, instead of the correlation coefficient $R^2_{\rm TRF}$ (Equation 2.47), we chose a correlation coefficient analogous to R^2_{xy}, defined in terms of an ensemble average of the ratios of tangent point slopes at different scales, this could result in terms that approach either zero or infinity, effectively under- or over-weighting those data points relative to the others in the ensemble. As we discussed in §2.5.1, this is precisely the situation that $R^2_{\rm TRF}$ was designed to handle, by defining the correlation of two fitted curves in terms of the absolute difference of their tangent point position angles, rather than the ratio of their tangent point slopes.

2.6 Model Comparison and Parameter Estimation

2.6.1 Comparing Models

There is a difference between *model comparison* and *parameter estimation*. Consider a hypothesis, or model, H described by M model parameters $\{\vartheta_m\}$. If this were the only possible model that could describe the observed data D and the prior information I, the probability of its being true is naturally equal to one. That is:

$$\int_{\vartheta_1} \cdots \int_{\vartheta_M} p(\vartheta_m|DI) d\vartheta_1 \cdots d\vartheta_M = 1\,, \tag{2.48}$$

where $p(\vartheta_m|DI)$ is the *posterior probability* of the model (Equation 2.1).

However, in reality there are multiple, or even an infinite number N of possible models $\{H_n\}$, each described by a set of M_n parameters $\{\vartheta_{mn}\}$, that might explain the

data and the priors. In this case, normalization of the posterior probability requires:

$$\sum_{n=1}^{N} \int_{\vartheta_{n1}} \cdots \int_{\vartheta_{nM_n}} p(\vartheta_{nm}|DI) d\vartheta_{n1} \cdots d\vartheta_{nM_n} = 1\,. \tag{2.49}$$

In practice, all possible models cannot be tested, or even known. The consequence of this fact is that the sum of the integrated posterior probabilities of the models that are being investigated is also an unknown value < 1. In other words, the *absolute* probabilities of a given model, or of a given parameterization of a model, cannot be known.

However, the probability of models, or parameterizations of models *relative* to one another can be investigated. For example, the probability of model H_1 relative to that of H_2 is given by the ratio:

$$\frac{\int_{\vartheta_{11}} \cdots \int_{\vartheta_{1M_1}} p(\vartheta_{1m}|DI) d\vartheta_{11} \cdots d\vartheta_{1M_n}}{\int_{\vartheta_{21}} \cdots \int_{\vartheta_{2M_2}} p(\vartheta_{2m}|DI) d\vartheta_{21} \cdots d\vartheta_{2M_2}}\,, \tag{2.50}$$

and the probability of model H_1 relative to a collection of models $\{H_2 \ldots H_N\}$ is:

$$\frac{\int_{\vartheta_{11}} \cdots \int_{\vartheta_{1M_1}} p(\vartheta_{1m}|DI) d\vartheta_{11} \cdots d\vartheta_{1M_n}}{\sum_{n=2}^{N} \int_{\vartheta_{n1}} \cdots \int_{\vartheta_{nM_n}} p(\vartheta_{nm}|DI) d\vartheta_{n1} \cdots d\vartheta_{nM_n}}\,. \tag{2.51}$$

For example, in the case of GRBs, we may wish to compare models based on fundamentally different assumptions about the physical source of the afterglow emission. The general consensus is that afterglows are due to synchrotron emission from shocks in ultra-relativistic jets generated in the final collapse of massive, rapidly rotating stars (the collapsar relativistic fireball/blast wave model). The standard forms of the blast wave model predict intrinsic afterglow emission that can be parameterized as a series of smoothly broken power laws in frequency and time. But other physical models for GRB afterglows and progenitors have been proposed (see, e.g., Mészáros, 2002; Woosley

& Bloom, 2006): Poynting flux dominated collapsar jets; pulsar jets; supernova "cannonballs"; neutron star-black hole mergers; supermassive black hole formation; and primordial black hole evaporation, to name a few. Such models may require a fundamentally different parameterization of the intrinsic emission from the smoothly broken power laws of the standard blast wave model.

In this dissertation, we do not compare GRB afterglow models with fundamentally different parameterizations. Rather, we introduce a single, highly flexible model, based on the consensus blast wave emission model, combined with line-of-sight extinction and absorption models that we develop in Chapter 3. We will, however, have occasion to compare different (sub)parameterizations of this model, and to compute uncertainties in the fitted model parameter values. However, one must always keep in mind that the model fitting results we (or for that matter, anyone) present are subject to the limitation that their absolute probabilities cannot be known.

2.6.2 Comparing Parameterizations of a Model

In this work, our approach is first to introduce a model that is sufficiently flexible to be able to describe most data sets for at least one, and usually multiple, degenerate parameterizations. If necessary, we can narrow our focus to different, physically interesting subsets of the parameter space. For example, consider the radial density profile $n(r)$ of the circumburst medium. In practice, this density profile is usually approximated by a simple power law, $n(r) \propto r^k$, although any density profile is mathematically (if not physically) possible. In the standard blast wave model, the temporal and spectral indices of the synchrotron emission in various regimes are functions of the circumburst density index k (Starling et al., 2008). However, even if we limit ourselves to a power-law profile, allowing k to be an entirely free parameter often results in an inability to break degeneracies between other, physically interesting fitted model parameters.

Instead, we often are forced to limit our investigations to the simple, but theoretically favored cases of a constant-density medium ($k = 0$) and a wind-blown medium ($k = -2$) (Chevalier & Li, 1999). Both the constant-density and wind-blown models are subsets, or *nested models*, of our more general model.

In theory, one can compare a nested model H' to its more general model H by evaluating the ratio:

$$\frac{\int_{\vartheta_1} \cdots \int_{\vartheta_M} p'(\vartheta_m|DI) d\vartheta_1 \cdots d\vartheta_M}{\int_{\vartheta_1} \cdots \int_{\vartheta_M} p(\vartheta_m|DI) d\vartheta_1 \cdots d\vartheta_M}, \tag{2.52}$$

where, in the case of the nested model, some of the parameters $\{\theta_m\}$ might be fixed, or written as functions of other model parameters. In practice, this requires numerical integration over M parameters with a sufficiently dense grid. However, in the models we investigate, M is sufficiently great to make this impractical (see §2.6.4). In lieu of this, we instead compare the best-fit total posterior probabilities of these models and evaluate how consistent or discrepant the nested model is with the general model, assuming a $\Delta\chi^2$ distribution. For example, if the nested model has one fewer free parameter, $(-2 \ln p')_{\min} - (-2 \ln p)_{\min} = 1$, 4 and 9 corresponds to the nested model being ruled out at the 1-, 2- and 3σ confidence levels, respectively.

2.6.3 Computing Best-Fit Parameters: *Galapagos*

To find a best-fit GRB afterglow model, it is necessary to explore a rather large parameter space. The full emission, absorption and extinction model contains well over 50 parameters. Most of these, however, are either constrained by priors on the parameter values themselves, or by empirical relationships by which a parameter is a function of one or more other parameters. These priors considerably reduce the volume of the parameter space we must explore. Our most general afterglow emission, extinction and absorption model consists of only 10-13 truly free parameters: 7-8 free emission parameters, plus 3-5 free parameters describing the source frame dust extinction and neutral

and molecular hydrogen absorption (see Chapter 3). Still, it is a daunting challenge to explore this parameter space completely and efficiently. The best tool we have found is the genetic algorithm, implemented in our model-fitting software package, *Galapagos*. We have found genetic algorithms to be much more computationally efficient at exploring large and complex parameter spaces, and at avoiding local minima, than other stochastic optimization schemes (e.g., Markov-Chain Monto-Carlo/simulated annealing heuristics), as well as being easily scalable to parallel-processor machines (Foster & Reichart 2011, in preparation).

The user initially provides *Galapagos* with the observed data, a set of priors, and an initial "search box", or guess of the numerical range that the value of each parameter may take. A single list of parameter values, or "genes", is called a "chromosome", and a given model parameterization, or "organism", consists of one or more chromosomes. *Galapagos* creates a set, or "population" of organisms, each of whose genes are selected randomly from the initial parameter search box ranges. The "fitness" (i.e., $-2 \ln \mathcal{L}$) of each organism is computed, and the population is sorted into a list of most- to least-fit. Then, some fraction (usually half) of the least fit organisms in the population is "killed off"; the remaining organisms are allowed to "breed", randomly exchanging genes, creating a new population of "offspring" organisms to replace those that were killed off. The process continues: each generation's organisms breed and their fitnesses are ranked, with the least fit being killed off and the most fit allowed to breed again, until the fitnesses of the organisms in the population differ from one another only by some small value – typically, this results in a population of identical organisms, or "clones". Since the true best-fit genes may not have been created in the original, randomized population – or, if they were, may not have survived the selection process – a new randomized population is generated, with search boxes recentered on the previous step's best-fit parameter values (the search box is either reduced or expanded by some

factor from its previous width, depending on how close the best-fit parameter value is to the center of the previous search box). The selection and breeding process begins again, until a new best-fit/clone population is reached. The entire process iterates until the fitness of successive best-fit/clone populations converges on the true global minimum of $-2\ln\mathcal{L}$, providing the best-fit parameter values.

It is often the case that a single set of model parameter values is inadequate to describe the entire range of the observed data. In the case of GRB afterglows, this most often occurs when emission or absorption characteristics change discontinuously with time (frequently during a gap between two nights' observations). For this reason, we refer to subsets of the data, together with their list of model parameter values, as "time slices" (though, generally speaking, the data can be subgrouped in any number of ways, according to criteria other than time). *Galapagos* has the flexibility to allow any number of time slices, and can constrain, or "link", certain parameters so that they always have the same value across all time slices (or among any arbitrary subset of time slices), while allowing other parameters to vary freely from slice to slice. In the genetic language introduced above, each time slice is a single chromosome in a multi-chromosome organism. For example, the redshift of the burst, $z_{\rm GRB}$, will always be linked across all time slices, as will parameters describing the Lyα forest absorption and Milky Way dust extinction along the burst's line of sight. But it may be that the burst experiences, say, a discontinuous re-brightening due to energy injection, or that absorbing material local to the burst is modified over the duration of the afterglow, in which case we may want to allow certain emission, extinction and absorption parameters to vary independently from slice to slice. For example, in the analysis of the afterglow of GRB 090313 in §4, we will explore the relative likelihood of models for which different subsets of the model parameters are linked between the first and second night of observation and find, in this case, that the observed discontinuity in

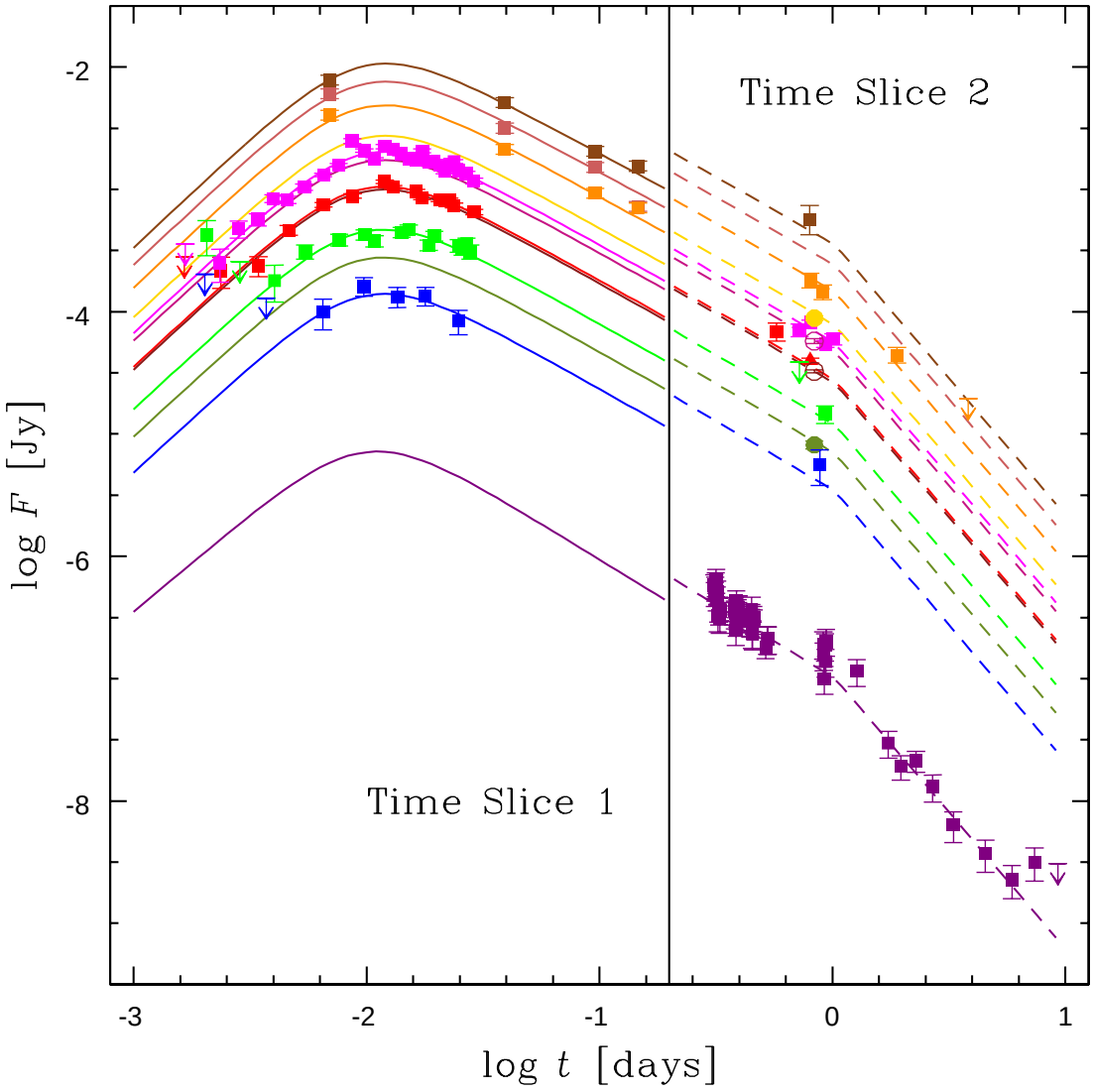

Figure 2.20 GRB 090313 best-fit light curve, with two time slices. The two time slices are divided at $\log t = -0.7$. Solid curves show model fits to data in the first time slice, dashed curves to data in the second time slice, for each photometric bandpass, from X-ray (violet) to NIR K-band (brown). All model parameters except for those describing intrinsic flux normalization are linked across both slices. See §4 for more details.

brightness is better explained by an intrinsic rebrightening than by variable extinction or absorption (Figure 2.20).

2.6.4 Computing Probability Distributions of Fitted Parameters

It is often assumed that the uncertainty in a fitted parameter is described by a $\Delta\chi^2$ distribution, as we had to do in the context of nested model comparisons in §2.6.2. However, in practice, this is not always the case, and may actually be a rarity.

The proper way to compute the probability distribution $p(\vartheta_i)$ of a fitted parameter ϑ_i is to compute the *marginalized probability distribution*. This requires marching through a dense grid over the entire M-dimensional parameter space, computing the posterior probability at each grid point, and then integrating over all dimensions except that of the parameter in question:

$$p^{\rm m}(\vartheta_i) = \int_{\vartheta_1} \cdots \int_{\vartheta_{i-1}} \int_{\vartheta_{i+1}} \cdots \int_{\vartheta_M} p(\vartheta_m|DI) d\vartheta_m \,. \tag{2.53}$$

This is computationally impractical for $M \gtrsim 5$.

There are two alternative approaches to estimating the probability distribution of a fitted parameter. A common, but wrong, approach is to compute the *cross-sectional probability distribution*, by holding all other fitted parameters fixed at their best-fit values $\hat{\vartheta}_m$, and "marching" in a one-dimensional grid for the parameter ϑ_i in question, computing the posterior probability at each step. That is:

$$p^{\rm c}(\vartheta_i) = p(\hat{\vartheta}_1, \ldots, \hat{\vartheta}_{i-1}, \vartheta_i, \hat{\vartheta}_{i+1}, \ldots, \hat{\vartheta}_M) \,. \tag{2.54}$$

If there is any correlation among the fitted parameters, this cross-sectional probability

distribution will be narrower than the marginal distribution, leading to underestimation of the error bars in the fitted parameters.

Instead, we compute the *projected probability distribution*, $p^{\mathrm{p}}(\vartheta_i)$, of each fitted parameter. This is done by marching along a one-dimensional grid for the parameter in question, holding the parameter fixed at each step and re-fitting with all the other parameters allowed to be free:

$$p^{\mathrm{p}}(\vartheta_i) = p(\hat{\vartheta}'_1(\vartheta_i), \ldots, \hat{\vartheta}'_{i-1}(\vartheta_i), \vartheta_i, \hat{\vartheta}'_{i+1}(\vartheta_i), \ldots, \hat{\vartheta}'_M(\vartheta_i)) \,. \tag{2.55}$$

If there is any correlation among the fitted parameters, this projected probability distribution will be broader than the marginalized distribution, leading to overestimation of the error bars in the fitted parameters. In this sense, projected probability distributions provide *conservative* error bars. Figure 2.21 illustrates the cross-sectional and projected confidence intervals for one parameter of a hypothetical, correlated two-parameter probability distribution. Figure 2.22 illustrates the resulting cross-sectional, marginalized and projected probability distributions. We note that in the case of parameters whose uncertainties are uncorrelated, all three approaches yield the same results. Correlations can often, but not always, be avoided by careful design of the model (see §2.6.6).

2.6.5 Computing Error Bars of Fitted Parameters

Having obtained a table of the projected probability distribution $p(\vartheta_i)$ of a model parameter ϑ_i, we are now in a position to compute projected confidence intervals, or error bars, on the best-fit parameter value. A 1-, 2- or 3σ confidence interval corresponds to a range of parameter values for which the integrated probability distribution is 68.2639%, 95.4500% or 99.7300% of the total integrated area under the distribution, respectively. Note that, strictly speaking, there are an infinite number of such ranges for each con-

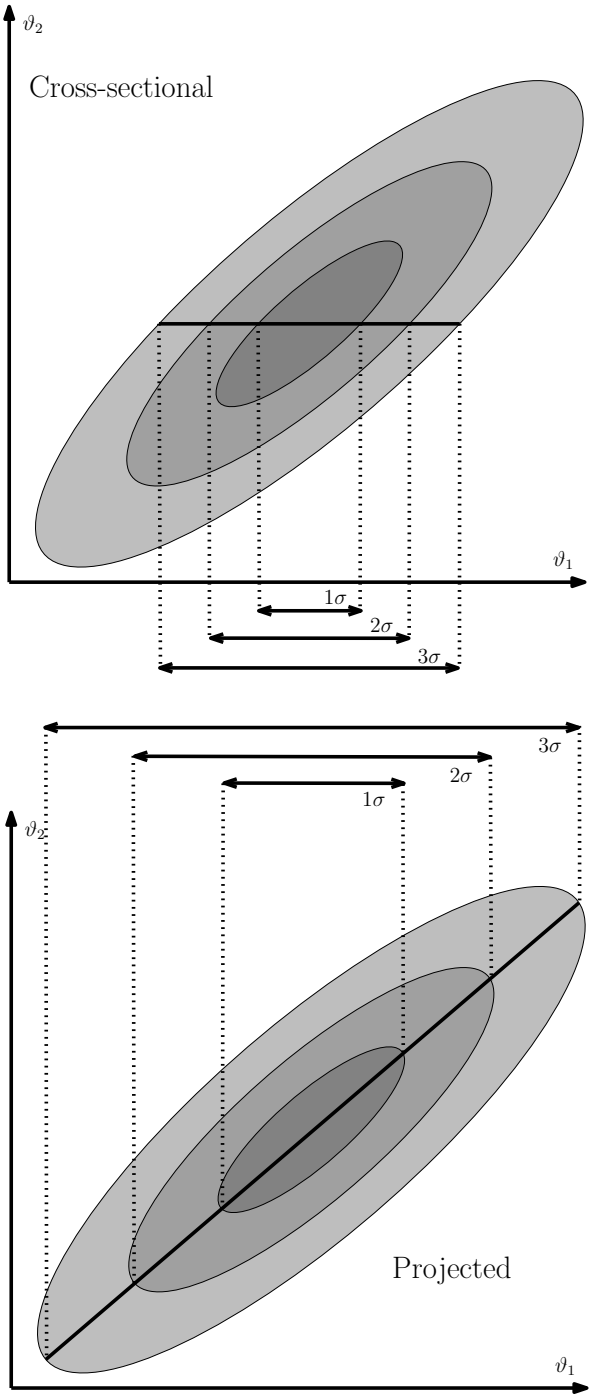

Figure 2.21 Illustration of cross-sectional and projected confidence intervals for a parameter ϑ_1, given a hypothetical, correlated two-parameter probability distribution $p(\vartheta_1, \vartheta_2)$. Shaded regions indicate the 1-, 2- and 3σ confidence regions of $p(\vartheta_1, \vartheta_2)$. Top panel: Cross-sectional confidence intervals. Bottom panel: Projected confidence intervals.

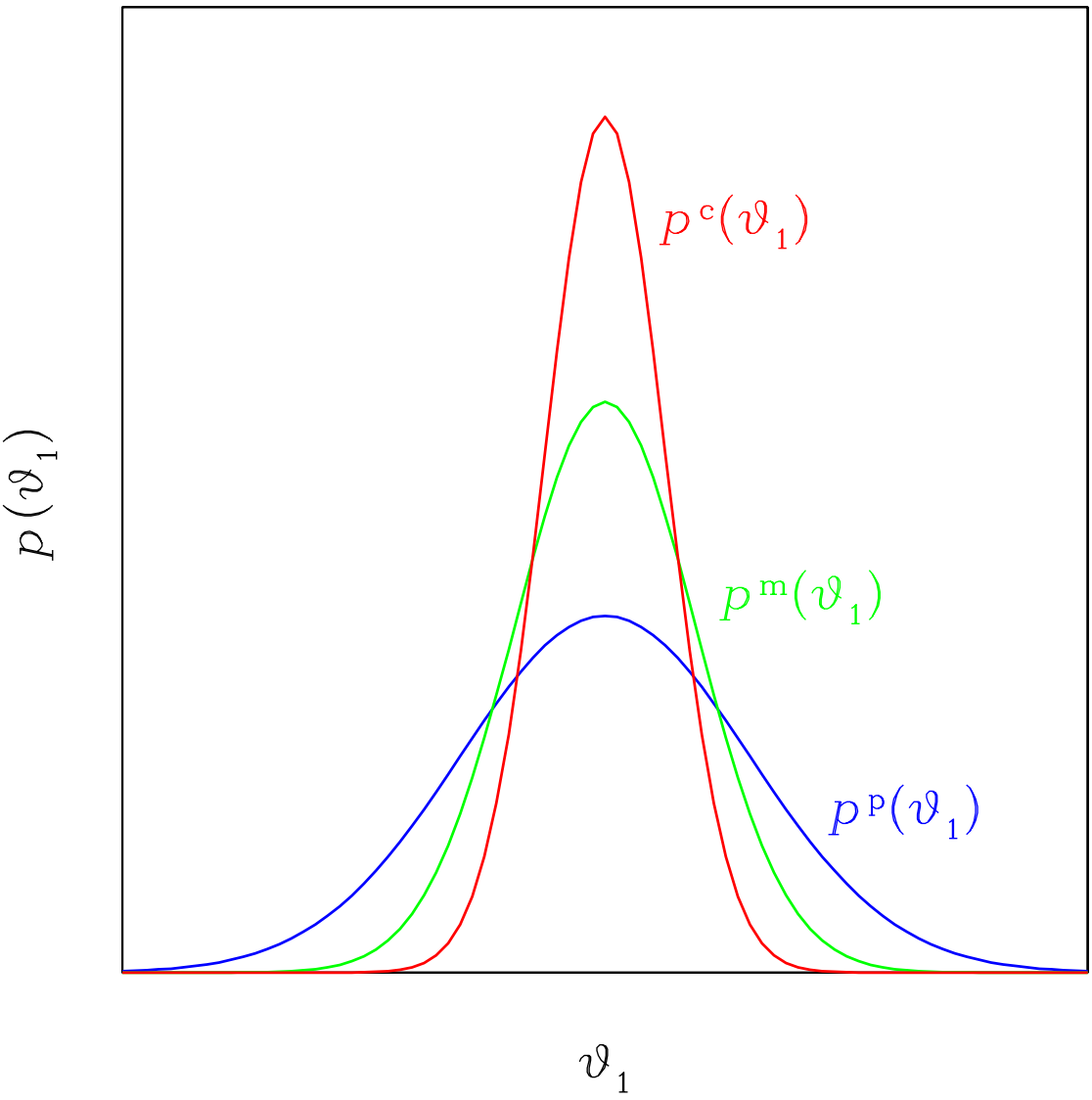

Figure 2.22 Comparison of cross-sectional ($p^{\mathrm{c}}(\vartheta_1)$, in red), marginalized ($p^{\mathrm{m}}(\vartheta_1)$, in green) and projected ($p^{\mathrm{p}}(\vartheta_1)$, in blue) probability distributions for a parameter ϑ_1, given the hypothetical two-parameter probability distribution $p(\vartheta_1, \vartheta_2)$ in Figure 2.21. If ϑ_1 and ϑ_2 are correlated, $p^{\mathrm{c}}(\vartheta_1)$ will be narrower than $p^{\mathrm{m}}(\vartheta_1)$, while $p^{\mathrm{p}}(\vartheta_1)$ will be broader.

fidence level, depending on where one chooses to place the limits of integration. What we are interested in are the *most compact* confidence intervals for a given probability distribution.

In practice, to obtain an accurate $N\sigma$ confidence interval, we typically tabulate projected $p(\vartheta_i)$ by marching in ϑ_i with a step size that gives ~ 100 entries between the values for which $-2\ln\mathcal{L} - (-2\ln\mathcal{L})_{\min} = (N+2)^2$ on either side of the best fit (so, to obtain accurate 1σ confidence intervals, we march until $-2\ln\mathcal{L} - (-2\ln\mathcal{L})_{\min} = 9$). The most compact confidence intervals are obtained by a technique we refer to as "lowering the bar". Figure 2.23 illustrates this concept, for a hypothetical bimodal probability distribution. Starting at the peak of the distribution, which corresponds to the best-fit parameter value, we lower an imaginary horizontal bar of constant $p(\vartheta_i)$ in small steps. At each step, we numerically interpolate the tabulated values of $p(\vartheta_i)$ to find the values of ϑ_i for which $p(\vartheta_i)$ is equal to the height of the horizontal bar, and numerically integrate the area under the probability distribution between those values. We adjust the height of the horizontal bar until the this area equals 68.2639%, 95.4500% and 99.7300% of the total integrated area under the tabulated probability distribution, which gives us the most compact 1-, 2- and 3σ confidence intervals of parameter ϑ_i.

Note that, as in the example of Figure 2.23, there are times when one or more of these most compact confidence intervals may be disjoint. In this example, there are, in fact, two such 1σ confidence intervals, each roughly centered on the two peaks in the bimodal distribution (however, the 2- and 3σ confidence intervals are contiguous). In such cases, one must be careful in reporting the best-fit parameter value and its uncertainty; strictly speaking, describing this distribution at the 1σ confidence level would require reporting two "best-fit" values of ϑ_i, each with its own error bars and relative weighted probability. In such cases, it is often better to present instead the actual probability distribution graphically, or in tabular form, with confidence intervals

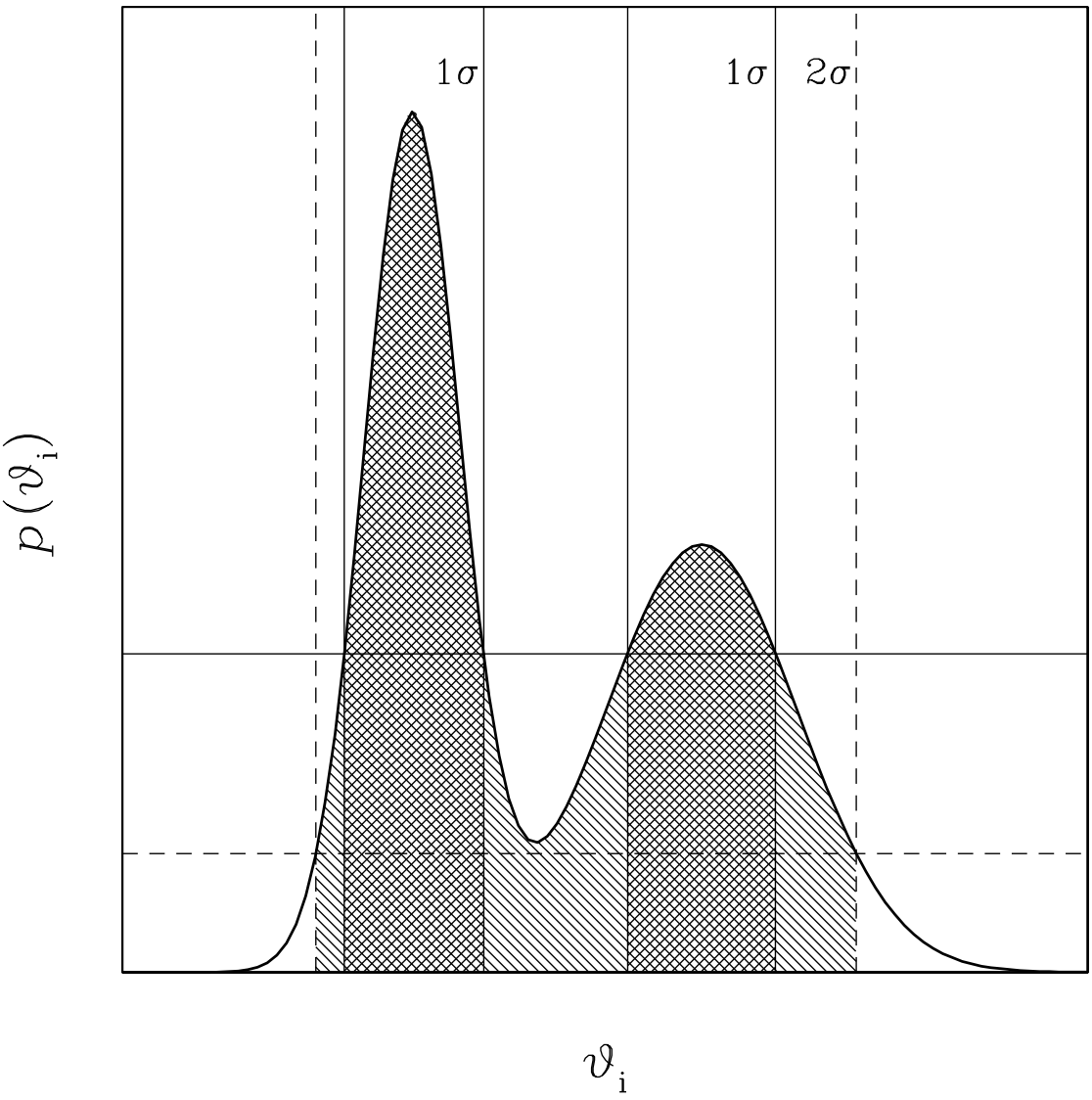

Figure 2.23 Illustration of confidence intervals for a bimodal fitted parameter probability distribution $p(\vartheta_i)$. Shaded regions show the most compact 1- and 2σ confidence intervals; the area of the shaded regions are 68.2689% and 95.4500% of the total area under the curve. This distribution has two, disjoint 1σ confidence intervals, but a single 2σ interval. The area of 3σ confidence region (not shown) would correspond to 99.7300% of the total area.

labeled as in Figure 2.23. If this probability distribution were intended to serve as a prior on ϑ_i in other model fits, we would most likely implement it by actually fitting a sum of two Gaussians model to the distribution. Fortunately, we find that in many cases, projected probability distributions fall monotonically from the best-fit peak, and can usually be well approximated by a Gaussian, or asymmetric Gaussian distribution (see, e.g., Figure 2.3 in §2.1). Utilizing such analytic approximations to prior probability distributions in our model-fitting algorithms can provide a significant advantage in numerical speed.

2.6.6 Correlated Uncertainties in Fitted Parameters and Pivot Points

We make efforts to minimize correlation among fitted parameters. For example, if we assume a linear model of the form $y(x) = mx + b$, the best-fit slope m will typically be correlated with the fitted intercept b. This is due entirely to the arbitrary choice of placing the intercept at $x = 0$; depending on the actual x-range of the data set, this choice can introduce a kind of lever arm effect whereby fitted m depends on b and vice versa. This correlation can be minimized by choosing a particular "pivot point", x_p, somewhere in the middle of the data range, so that $y(x) = m(x - x_p) + b_p$.

In practice, we find the optimum pivot point by marching in m in a fine grid very close to the best-fit, refitting to b at each step. Consider two such linear fits, with best-fit parameters (b_1, m_1) and (b_2, m_2) (Figure 2.24). The pivot point x_p corresponds to the value of x at which these two lines intersect:

$$\begin{aligned} m_1 x_p + b_1 &= m_2 x_p + b_2 \\ x_p &= -\frac{b_1 - b_2}{m_1 - m_2} \equiv -\mu \,. \end{aligned} \tag{2.56}$$

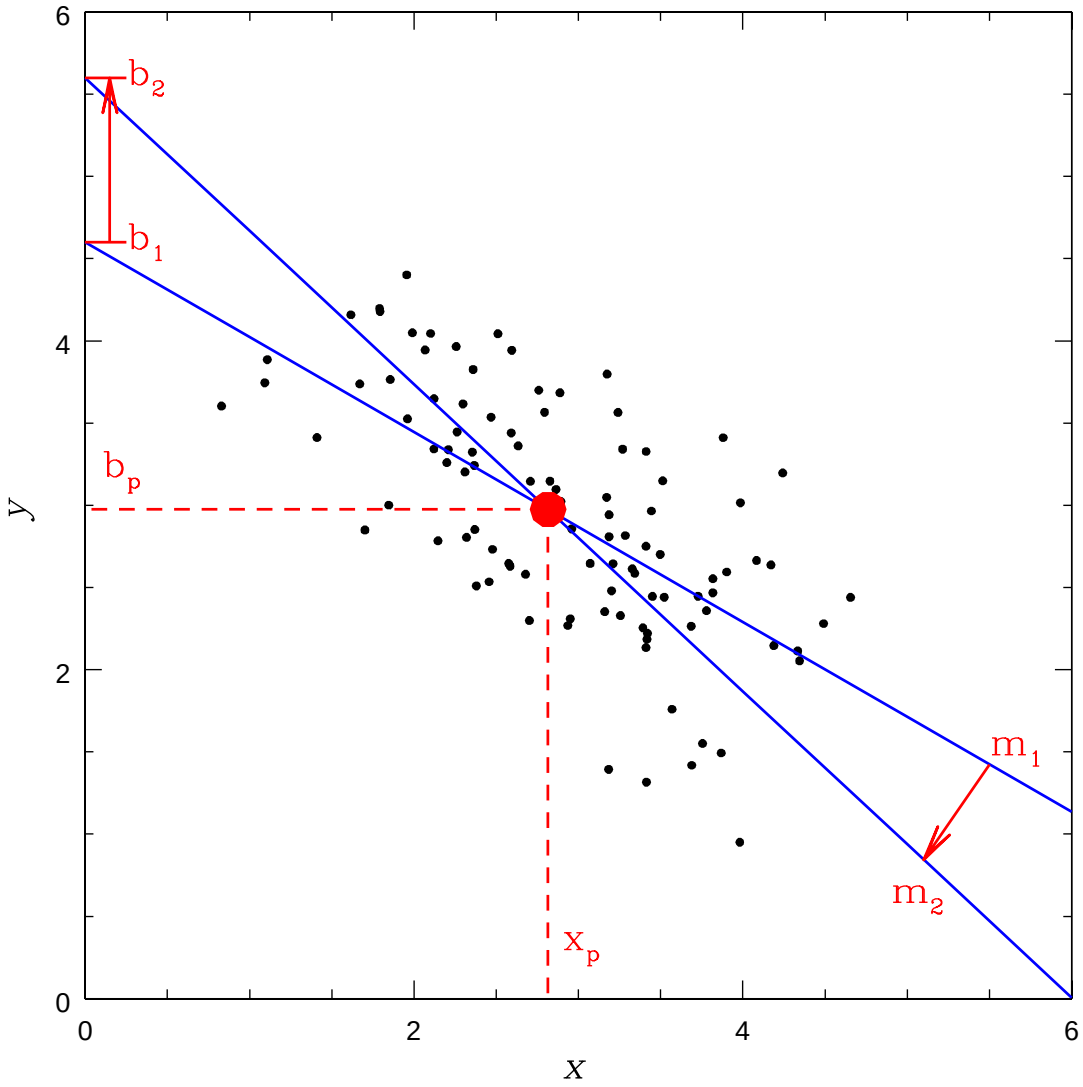

Figure 2.24 Illustration of linear pivot point determination. The two lines $y = m_1x+b_1$ and $y = m_2x + b_2$ are fits that might be obtained in an error bar march in the slope parameter m. Clearly, any uncertainty in m will be correlated with uncertainty in the intercept b, and vice versa. If we instead fit to $y = m(x - x_p) + b_p$, where the pivot point $x_p = -(b_1 - b_2)/(m_1 - m_2)$, correlations between uncertainties in m and b_p will be minimized.

By extension, when we plot the best-fit b_i vs. m_i for all the fits in the march, we find a clear linear correlation, with slope μ. The best choice of pivot point, to minimize the correlation between slope and intercept, is $x_p = -\mu$, so that $b_p = b - m\mu$. We then conduct the error bar marches with the model $y(x) = m(x - x_p) + b_p$ (holding x_p fixed) to determine projected probability distributions for m and b_p.

For more complicated, non-linear models, we can only attempt to parameterize them, and choose a coordinate basis, in such a way that certain segments are asymptotically linear, and analyze independently the fitted model parameters that describe these segments to find their individual pivot points. In practice, this minimizes the correlation between slope and intercept of each approximately linear segment, though it does not address the issue of cross-correlation among the multiple segments, or correlations among parameters that describe non-linear regions of the curve.

2.7 Advanced Topics

2.7.1 Choice of Basis and Basis Transformations

In all the preceding discussion of fitting model distributions to two-dimensional data sets, we have assumed that both the intrinsic uncertainties (error bars) and the extrinsic scatter (slop) are normally distributed in the space in which we conduct the fit. This assumption considerably simplifies the resulting statistics, allowing us to approximate the various probability integrals analytically. However, there are times when either the intrinsic uncertainty or the extrinsic scatter of the data in one or both dimensions are *not* both well described by a symmetric, Gaussian probability distribution function in the same space, or basis. For example, we measure the flux density, F_ν, of a GRB afterglow using a CCD camera, which is an intrinsically linear flux density-measuring instrument. The flux density is measured in the form of photon counts in each CCD

pixel element, and the uncertainty in the flux density follows a Poisson distribution, which approaches a Gaussian distribution in the limit of count numbers $N \gtrsim 10$, so the intrinsic uncertainty in our measurements, to a good approximation, has a symmetric Gaussian probability distribution in linear F_ν space. However, the calibration of our flux density measurements using standard stars is done in magnitude space, and uncertainties in that calibration are assumed to be normally distributed in magnitude, or $\log F_\nu$, space. Furthermore, the unmodeled physical processes that contribute to the extrinsic scatter (slop) in the observed flux of a GRB afterglow are typically assumed to have a normal distribution in $\log F_\nu$. So, while the data have normal intrinsic uncertainties in linear F_ν, the model distribution is normally distributed in $\log F_\nu$; since we must choose one or the other basis in which to conduct the fit, we must transform one or the other of the distributions into that space.

As we discuss in some detail in §3.5, we choose to conduct our GRB model fits in linear F_ν space, using a simple, one-dimensional χ^2-like statistic. This requires that we transform the log-normal model distribution into that linear space. Besides being in line with our overall philosophy of "bringing the model to the data", this approach also allows us to easily include non-detections (typically reported as upper limits in $\log F_\nu$ or magnitude space) in our data set; an upper limit in $\log F_\nu$ may be approximated as a zero-mean Gaussian in linear F_ν (See §2.7.3).

Most generally, if the probability distribution function in a particular basis x has the functional form $p(x)$, then in the transformed basis $f(x)$, conservation of probability requires that the transformed distribution $p(f(x))$ takes the form:

$$\begin{aligned} p(f(x))df(x) &= p(x)dx \\ p(f(x)) &= p(x)\left(\frac{df(x)}{dx}\right)^{-1}. \end{aligned} \tag{2.57}$$

So, if the GRB flux model distribution in $\log F_\nu$ space has the log-normal form:

$$p(\log F_\nu) = G(\log F_\nu; \log F_\nu^{\rm mod}, \sigma_{\log F_\nu}) \,, \tag{2.58}$$

then, in linear F_ν space, $p(F_\nu)$ will take the form:

$$\begin{aligned} p(F_\nu) &= G(\log F_\nu; \log F_\nu^{\rm mod}, \sigma_{\log F_\nu}) \left(\frac{d \log F_\nu}{d F_\nu} \right) \\ &= G(\log F_\nu; \log F_\nu^{\rm mod}, \sigma_{\log F_\nu}) \left(\frac{1}{F_\nu \ln 10} \right) \,. \end{aligned} \tag{2.59}$$

In §2.7.2 we show that the pdf in Equation 2.59 can be approximated by an asymmetric Gaussian function.

2.7.2 Asymmetric Gaussian Probability Distributions

It can be shown analytically that, in linear F_ν space, the transformed log-normal $p(\log F_\nu)$ of Equation 2.59 is a maximum at:

$$F_\nu^{\rm max} = 10^{(\log F_\nu^{\rm mod} - \sigma^2_{\log F_\nu} \ln 10)} \,. \tag{2.60}$$

If the slop $\sigma_{\log F_\nu} \lesssim 0.1$ (which, we find, is almost always the case for GRB afterglows), $p(F_\nu)$ can be well approximated as a *skew-normal*, or *asymmetric Gaussian* function that peaks at $F_\nu^{\rm max}$ (Equation 2.60):

$$p(F_\nu) \approx G_A(F_\nu; F_\nu^{\rm max}, \sigma_{F_\nu \pm}) \,, \tag{2.61}$$

where G_A is the normalized asymmetric Gaussian function:

$$G_A(F_\nu; F_\nu^{\max}, \sigma_{F_\nu\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{F_\nu+} + \sigma_{F_\nu-})} \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{F_\nu - F_\nu^{\max}}{\sigma_{F_\nu+}}\right)^2\right] & \text{if } F_\nu \geq F_\nu^{\max} \\ \exp\left[-\frac{1}{2}\left(\frac{F_\nu - F_\nu^{\max}}{\sigma_{F_\nu-}}\right)^2\right] & \text{if } F_\nu < F_\nu^{\max} \end{cases} , \tag{2.62}$$

and where the 1σ widths of the asymmetric Gaussian are given by the empirical formulas:

$$\begin{aligned} \sigma_{F_\nu+} &\approx F_\nu^{\max}(2.3029\sigma_{\log F_\nu} + 2.6293\sigma_{\log F_\nu}^2 - 3.6945\sigma_{\log F_\nu}^3) \\ \sigma_{F_\nu-} &\approx F_\nu^{\max}(2.3027\sigma_{\log F_\nu} - 2.6544\sigma_{\log F_\nu}^2 - 4.0699\sigma_{\log F_\nu}^3)\,. \end{aligned} \tag{2.63}$$

We emphasize that these are, indeed, *approximations*, and only valid for cases where the slop parameter $\sigma_{\log F_\nu} \lesssim 0.1$. In general, the transformed distribution's peak, and the most compact 1σ confidence interval around that peak must be found numerically, using techniques like that presented in §2.6.5. In some cases, the transformed probability distribution may then be expressed in terms of an asymmetric Gaussian distribution with a peak and widths that reflect that confidence interval. Other functional transformations of basis may yield distributions that are not well approximated by asymmetric Gaussian distributions, and must be treated with care. Strictly speaking, the proper approach to dealing with generalized transformed, non-normal distributions is to evaluate the probability integrals numerically; unfortunately, this is prohibitively computationally expensive.

In cases where the model distribution, or a data point's intrinsic probability distribution, or both can be approximated by asymmetric Gaussian functions, we have formulated an empirical approximation of the joint probability. The expression for the joint probability of data point n with the model distribution (Equation 2.6) takes the

form:

$$p_n(\vartheta_m, \sigma_{x\pm}, \sigma_{y\pm} | x_n, y_n, \sigma_{x,n\pm}, \sigma_{y,n\pm}) = \int_{x'} \int_{y'} \int_{u_n} \int_{v_n} dv_n du_n dy' dx' f(x', y') g(x, y) \delta(v_n - v_{c,n}(u_n; \vartheta_m)) \times G_A(x', x, \sigma_{x\pm}) G_A(y', y, \sigma_{y\pm}) G_A(x', x_n, \sigma_{x,n\pm}) G_A(y', y_n, \sigma_{y,n\pm}) \,. \quad (2.64)$$

The x' and y' integrals can be shown to be equivalent to convolutions, i.e.:

$$\int_{x'} G_A(x', x, \sigma_{x\pm}) G_A(x', x_n, \sigma_{x,n\pm}) dx' \equiv G_A(x, 0, \sigma_{x\mp}) * G_A(x, x_n, \sigma_{x,n\pm}) \,. \quad (2.65)$$

In Appendix A, we describe an approximation of the convolution of two asymmetric Gaussian functions as a single asymmetric Gaussian:

$$G_A(x, 0, \sigma_{x\mp}) * G_A(x, x_n, \sigma_{x,n\pm}) \approx G_A(x, x_n + \delta_{x,n}, \Sigma_{x,n\pm}) \,, \quad (2.66)$$

where

$$\Sigma^2_{x,n\pm} \equiv \sigma^2_{x\mp} + \sigma^2_{x,n\pm} \,, \quad (2.67)$$

and where $\delta_{x,n}$ is an empirical function of $\sigma_{x\mp}$ and $\sigma_{x,n\pm}$. In other words, the x' and y' integrals in the joint probability are approximately equivalent to shifting the data point centroid to $(x_n + \delta_{x,n}, y_n + \delta_{y,n})$, and adding the slops $(\sigma_{x\mp}, \sigma_{y\mp})$ to the intrinsic error bars $(\sigma_{x,n\pm}, \sigma_{y,n\pm})$ in quadrature.

For the one-dimensional, χ^2-like statistic we employ to fit light curves to GRB flux measurements, the joint probability then takes the simple form:

$$p_n \approx \frac{2}{\sqrt{2\pi}(\Sigma_{F_\nu,n+} + \Sigma_{F_\nu,n-})} \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{F_{\nu,n} + \delta_{F_{\nu,n}} - F_\nu^{\max}}{\Sigma_{F_{\nu,n}+}}\right)^2\right] & \text{if } F_{\nu,n} + \delta_{F_{\nu,n}} \geq F_\nu^{\max} \\ \exp\left[-\frac{1}{2}\left(\frac{F_{\nu,n} + \delta_{F_{\nu,n}} - F_\nu^{\max}}{\Sigma_{F_{\nu,n}-}}\right)^2\right] & \text{if } F_{\nu,n} + \delta_{F_{\nu,n}} < F_\nu^{\max} \end{cases} , \quad (2.68)$$

where $F_\nu^{\max}$ is the peak of the transformed log-normal model distribution (Equation 2.60).

For two-dimensional statistics, like TRF, evaluating the joint probability of an asymmetric model distribution with the asymmetric intrinsic uncertainty of a data point is somewhat more complicated. We proceed, as before, by approximating the model curve $y_c(x;\vartheta_m)$ by a line tangent to the asymmetric convolved (and shifted) error "ellipse", taking care to keep track of which quadrant the tangent point lies in. However, since the tangent line will, in general, cross through three quadrants of the asymmetric convolved ellipse, the joint probability integral over the range $[-\infty,\infty]$ of Equation 2.14 must be broken into three segments. We discuss this procedure in Appendix B.

2.7.3 Treatment of Detections and Upper Limits in $\log F_\nu$ Space

In our own GRB afterglow photometry, flux density measurements are reported in linear F_ν space, with symmetric Gaussian error bars. But we often need to plot these measurements in $\log F_\nu$ (or, perhaps, magnitude) space. If the intrinsic pdf of a flux density measurement has the symmetric Gaussian form $G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}})$ in linear F_ν space, then, by Equation 2.57:

$$\begin{aligned} p(\log F_\nu) &= G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}}) \left(\frac{d\log F_\nu}{dF_\nu}\right)^{-1} \\ &= G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}}) \left(F_\nu \ln 10\right) . \end{aligned} \tag{2.69}$$

It can be shown analytically that the pdf of Equation 2.69 is a maximum at:

$$\log F_{\nu,n}^{\max} = \log\left[\frac{1}{2}\left(F_{\nu,n} + \sqrt{F_{\nu,n}^2 + 4\sigma_{F_{\nu,n}}^2}\right)\right] . \tag{2.70}$$

If a measurement $F_{\nu,n}$ has a sufficiently high signal-to-noise ratio, $p(F_\nu > 0) \geq 0.9973$, or:

$$\int_0^\infty G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}}) dF_\nu \geq 0.9973\,, \tag{2.71}$$

then we transform the distribution using Equation 2.69. In other words, if the area under the normalized Gaussian for $F_\nu > 0$ is equal to or greater than the area of the 3σ confidence interval, we treat the data point as detection. The transformed data point is plotted with a centroid value $\log F_{\nu,n}^{\max}$, given by Equation 2.70, and with asymmetric (non-Gaussian) error bars that must be determined numerically, using the "bar-lowering" technique of §2.6.5.

Measurements with low signal-to-noise ratios, $p(F_\nu > 0) < 0.9973$, or:

$$\int_0^\infty G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}}) dF_\nu < 0.9973\,, \tag{2.72}$$

are plotted as upper limits in $\log F_\nu$ space. The 3σ upper limit $\log F_{\nu,n}^{3\sigma}$ is the value for which:

$$\int_{-\infty}^{F_{\nu,n}^{3\sigma}} G(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}}) dF_\nu = 0.9973\,. \tag{2.73}$$

Similarly, the 2σ and 1σ upper limits, $\log F_{\nu,n}^{2\sigma}$ and $\log F_{\nu,n}^{1\sigma}$, are those for which the integral in Equation 2.73 equals 0.9545 and 0.6827, respectively. Note, however, that by this definition, the 1-, 2- or 3σ limits may not exist in $\log F_\nu$ space. For instance, if $F_{\nu,n}^{1\sigma} < 0$, then $\log F_{\nu,n}^{1\sigma}$ is undefined. (In the rare case that even the 3σ upper limit is undefined, it may only be possible to plot that data point as a 4σ, or higher, limit in $\log F_\nu$ space.)

Again, we only transform our flux measurements into $\log F_\nu$ space for the purposes of data visualization. All GRB afterglow fits are conducted in linear F_ν space, using the approximation of the transformed log-normal model distribution discussed above, in §2.7.2. We chose this approach precisely because it allows us to include low signal-

to-noise measurements, and even non-detections, in our data sets. In $\log F_\nu$ space, the intrinsic probability distributions of these measurements are highly asymmetric, and impossible to treat analytically – even leaving aside the fact that any fraction of the distribution lying below $F_\nu = 0$ is "thrown away" in the transformation from linear to log space.

We often are presented with the reverse scenario: measurements quoted as upper limits in $\log F_\nu$ (or magnitude) space in the literature. Most often, a 3σ upper limit $\log F^{3\sigma}_{\nu,n}$ is provided. In the absence of any other information about how they were computed, the best we can do is to assume that these upper limits correspond to zero-mean Gaussian distributions in linear F_ν space. The 1σ width of that distribution is the value of $\sigma_{F_{\nu,n}}$ for which:

$$\int_{-\infty}^{F^{3\sigma}_{\nu,n}} G(F_\nu; 0, \sigma_{F_{\nu,n}}) dF_\nu = 0.9973\,. \tag{2.74}$$

By this definition, $\sigma_{F_{\nu,n}} = F^{3\sigma}_{\nu,n}/2.7822$. Analytically, if 2- or 1σ upper limits are given, then $\sigma_{F_{\nu,n}} = F^{2\sigma}_{\nu,n}/1.6901$ or $F^{1\sigma}_{\nu,n}/0.4752$, respectively.[8]

[8] A common, but technically incorrect, practice is simply to assume that $\sigma_{F_{\nu,n}} = F^{3\sigma}_{\nu,n}/3$ (or $F^{2\sigma}_{\nu,n}/2$ or $F^{1\sigma}_{\nu,n}/1$). However, we err on the side of caution and, given a 3σ upper limit, employ the more conservative estimation, given by Equation 2.74, for the width of the zero-mean Gaussian pdf in linear F_ν space.

Chapter 3

Model

"You boil it in sawdust: you salt it in glue:
You condense it with locusts and tape:
Still keeping one principal object in view—
To preserve its symmetrical shape."
— Lewis Carroll, *The Hunting of the Snark*, Fit the Fifth

3.1 Modeling GRB Afterglows: An Overview

We are now in a position to begin constructing a comprehensive model that describes the observed flux density of a GRB afterglow as a function of frequency and time. We wish to specify $F_\nu(t;\vartheta_m)$ in Janskys (1 Jy = 10^{-23} erg s^{-1}cm^{-2}Hz^{-1}), where ν is the frequency of radiation detected in the observer's rest frame, t is the time since the onset of the GRB in the observer's frame (where $t = 0$ corresponds to the time of initial detection of gamma-ray emission by satellite detectors), and $\{\vartheta_m\}$ is a set of model parameters describing both the intrinsic emission and all sources of extinction and absorption along the line of sight to the burst.

We assume that the intrinsic emission of an afterglow is due to synchrotron radiation from shocks formed when a relativistic jet emerging from the GRB progenitor collides with, and is decelerated by, the circumburst medium (e.g., Mészáros & Rees, 1997;

Sari et al., 1998; Piran, 1999; Chevalier & Li, 1999; Granot et al., 2000; Mészáros, 2002, and references therein). This standard blast wave model, which we present in §3.2, predicts intrinsic emission that is described by smoothly broken power laws in frequency and time. From fitted spectral and temporal power-law indices of the intrinsic afterglow emission, we can infer both the energy distribution of the shocked electrons and the density profile of the circumburst medium. If our data are of sufficient time and frequency coverage, from the fitted frequencies of various spectral breaks and the overall flux normalization we can infer: the isotropic-equivalent energy of the blast wave, $E_{\rm iso}$; the fraction of that energy that is in electrons and in the magnetic field, ϵ_e and ϵ_B; and the density normalization of the circumburst medium (e.g., n for a constant-density medium, or A_* for a wind-driven medium).

However, the values of, and uncertainties in, the fitted intrinsic emission model parameters and any physical quantities inferred from them will be in error if sources of extinction and absorption are not also – and properly – modeled. Furthermore, the fitted values of the extinction parameters are of physical interest in their own right, particularly those that describe extinction and absorption in the circumburst environment. In the remainder of Chapter 3, we develop models for sources of line-of-sight extinction and absorption of the intrinsic GRB afterglow emission as a function of observing frequency ν. Our models include:

- Extinction due to dust in the Milky Way (§§3.3.1 & 3.3.2).
- Extinction due to dust in the GRB source frame, including both the circumburst environment and the host galaxy (§§3.3.3 & 3.3.4).
- Absorption due to neutral hydrogen in the GRB source frame (damped Lyα absorber/Lyman limit; §3.4.1).
- Absorption due to neutral hydrogen in the intergalactic medium (IGM) (Lyα

forest/Gunn-Peterson trough; §3.4.2).

It is important to note that these absorption and extinction models are very broadly applicable, not only to GRB afterglows, but also to any extragalactic point source of radiation (but the extinction model should not be used to model extended sources, such as galaxies, whose sources of emission are distributed throughout a dusty medium).

Our afterglow extinction and absorption models are described by dozens of parameters; however, the majority of them are constrained by priors. These prior constraints come in two types:

1. Prior constraints on secondary parameters describing the probability distribution of a single model parameter; and

2. Prior constraints on secondary parameters describing a model distribution fit to the correlation between two model parameters.

As an example of the first type of prior constraints, consider the Milky Way extinction parameter $R_V^{\rm MW}$ (§2.1). A typical value for this parameter is $R_V^{\rm MW} = 3.1$, and it is common practice to hold this parameter fixed at that value in extinction fits. But, in fact, $R_V^{\rm MW}$ is observed to have a distribution of values when measured along multiple lines of sight (see Figure 3.2). If we hold $R_V^{\rm MW}$ fixed at 3.1, the probability distribution of any other model parameter that we fit to will be overly narrow and, consequently, the uncertainty in its inferred value will be underestimated (§2.1). On the other hand, if we simply allow $R_V^{\rm MW}$ to be a free parameter (i.e., a flat prior), the probability distribution of any other model parameter that we fit to will be overly broad and, consequently, the uncertainty in its inferred value will be overestimated (§2.1). The proper approach, in Bayesian analysis, is to allow the parameter $R_V^{\rm MW}$ to vary, but subject to the constraint of some prior probability distribution $p(R_V^{\rm MW})$. In practice, we arrive at this prior by fitting a model to the distribution of prior observations of $R_V^{\rm MW}$; the model distribution

is, itself, described by a set of secondary parameters $\{\vartheta_m\}$, all of which are themselves constrained by priors $p(\vartheta_m)$, based on their fitted probability distributions. The purpose of these secondary priors is to quantify our uncertainty in the determination of the prior probability distribution $p(R_V^{\rm MW})$. In §3.3.2, we describe our formulation and implementation of the $R_V^{\rm MW}$ probability distribution model priors.

As an example of the second type of prior constraints, consider the observed correlation between the UV extinction parameters c_1 and c_2 (§2.1). The observed values of c_1 vs. c_2 along various lines of sight follow a linear trend, but with additional, unmodeled extrinsic scatter, or "slop", potentially in both the x- and y dimensions, about any best-fit line that is greater than can be accounted for by the intrinsic uncertainties in the data points alone (see Figure 3.8). In such a case, it is necessary to fit a *model distribution* to the 2D data set, with the slops (σ_x, σ_y) as free model parameters. This is precisely the problem that the TRF statistic, which we presented in Chapter 2, was designed for. The fitted model distribution is described by a curve, $y_c(x; \vartheta_m)$, convolved with a 2D Gaussian function with widths (σ_x, σ_y). In the case of the linear fit to c_1 vs. c_2, the model curve is defined by an intercept secondary parameter b^{c_1} and a position angle secondary parameter θ^{c_1}, both of which are constrained by priors $p(b^{c_1})$ and $p(\theta^{c_1})$ based on their fitted probability distributions. The model distribution is defined by this curve convolved with the 2D Gaussian slop $(\sigma_{c_2}^{c_1}, \sigma_{c_1})$. The extinction parameter c_1 is then a function of the (free) extinction parameter c_2, and of these secondary parameters. In §3.3.4.1, we describe our formulation and implementation of the c_1 vs. c_2 correlative model distribution priors.

In §3.5, we describe our GRB afterglow model fitting procedure, beginning with a general discussion of good and bad modeling practices. Our combined emission, extinction and absorption models describe the observed flux density of a GRB afterglow as a function of frequency and time, $F_\nu(t; \vartheta_m)$. To compare our model with the observed

photometric flux densities in a data set, we integrate the model spectral flux density F_ν over each data point's filter response function to obtain a filter response-weighted mean flux density; this is particularly important for modeling photometric observations at UV frequencies (in the GRB source frame), where our model includes spectral features like the 2175 Å dust extinction feature (the so-called "UV bump"), damped Lyα absorption, the Lyα forest and Lyman limit absorption. We model extrinsic scatter, or slop, as normally distributed in $\log F_\nu$ space, with the slop $\sigma_{\log F_\nu}$ fit to as a free parameter.[1] We transform the log-normal model distribution of the light curve into linear F_ν space and employ a 1D χ^2-like statistic to compute the likelihood function (§§2.7.1 & 2.7.2).

While computing the likelihood function for a given GRB model parameterization and photometric data set is computationally intensive, our genetic algorithm software, *Galapagos* (Foster & Reichart, in preparation; §2.6.3), makes it possible to explore the parameter space and arrive at a best fit in a reasonable amount of time (particularly when implemented on parallel-processor machines). *Galapagos* has also been designed with a number of highly flexible features that facilitate other aspects of our model-fitting procedure. Each data point can be assigned into one (or more) *calibration groups* (typically, one group per photometric filter per independently prepared data set). Our model then allows for *calibration offset parameters*, each constrained by a zero-mean Gaussian prior whose width is given by the measured (if known) or otherwise estimated calibration uncertainty. This allows such light curves to be shifted up or down by some amount Δ_c to correct for systematic calibration errors in each group (see §3.5). The

[1] Any unmodeled physical processes that might produce slop in the observed data are expected, a priori, to be in the F_ν dimension, and not in the t dimension. Slop in the t dimension would have to be due to unmodeled physical processes that introduced stochastic time delays in the light curve of the afterglow; while not impossible, we consider these likely to be negligible when compared to the multitude of plausible stochastic physical processes that might introduce unmodeled scatter in flux. Similarly, the intrinsic uncertainty in the time t of an observation is negligible when compared to the intrinsic uncertainty of a given photometric flux measurement. In the limiting case of effectively zero extrinsic and intrinsic scatter in, in this case, the t-dimension, the TRF statistic reduces to a 1D χ^2-like statistic (§2.4.2).

data can also be organized into separate *data groups* based on other criteria; often, in GRB fits, these data groups correspond to different time ranges, or "time slices", in the observation. With *Galapagos*, we can allow the value of a given model parameter to vary independently for each data group, or require that its value be *linked* across one or more data groups. For example, in fitting to a discontinuous GRB light curve, with parameter linking across time slice data groups we can determine the relative likelihood that the discontinuity is due to variable emission versus variable extinction. Parameters may also be *functionally linked*, so that the value of one may be a function of the values of one or more of the others; for instance, in the standard relativistic blast wave GRB emission model, the spectral and temporal indices β and α in different regimes can be expressed as various functions of the electron energy distribution index p and the circumburst density profile index k (see §3.2). This functional linking allows us to explore the relative likelihoods of *nested models*, or sub-parameterizations of a given model (§2.6.2). Finally, *Galapagos* can perform *error bar marches*, holding a parameter's value fixed at stepped intervals about its best-fit value, while allowing all other parameters to vary to find that parameter's best-fit *projected probability distribution* (§2.6.4).

3.2 GRB Emission Model

3.2.1 Standard Blast Wave Model

We model the intrinsic emission of a GRB afterglow using the standard ultrarelativistic blast wave model (e.g., Mészáros & Rees, 1997; Sari et al., 1998; Piran, 1999; Chevalier & Li, 1999; Granot et al., 2000; Mészáros, 2002, and references therein). In this model, the jet that emerges from the accretion disk around the newly formed black hole is modeled as a section of an expanding spherical shell. The afterglow is due to synchrotron emission from accelerated electrons in the relativistic shock formed as the leading edge

of the ejected material collides with, and is decelerated by, the circumburst medium. The energy of the accelerated electrons in the shock is assumed to have a power-law distribution:

$$\begin{aligned} N_e(E) &\propto E^p \qquad \text{if } E \geq E_m \,, \\ &= 0 \qquad \text{if } E < E_m \,, \end{aligned} \tag{3.1}$$

with an electron energy index $p < -2$ and a cutoff in the distribution at a minimum energy E_m to keep the total energy finite.

Figure 3.1 illustrates a typical theoretical spectral energy distribution for synchrotron emission for a GRB afterglow with such an electron distribution. The majority of electrons in the distribution have energy $E \sim E_m$. Therefore, the observed spectral energy distribution peaks at a frequency ν_m, which is the characteristic synchrotron frequency of an electron with energy E_m. The spectrum varies as $\nu^{1/3}$ for $\nu < \nu_m$, and as $\nu^{(p+1)/2}$ for $\nu > \nu_m$ (e.g., Rybicki & Lightman, 1979). If the blast wave is adiabatic, ν_m decreases with time as $t^{-3/2}$ (e.g., Mészáros & Rees, 1997). In typical GRB afterglows, ν_m passes through the optical/NIR region of the spectrum in the first few seconds, passing quickly into the far infrared/sub-millimeter region of the spectrum; hence, this peak is rarely observed in optical/NIR afterglow measurements.

Obviously, as the accelerated electrons continue to radiate, they lose energy. The power radiated by an electron of energy E is proportional to E^2. An electron of energy E will thus lose its energy to radiation on a timescale $t \propto E^{-1}$. In other words, at a given time t, any electrons with energy $E > E_c(t) = C(t)/t$ will be lost from the power-law distribution, where $C(t)$ is a function of the bulk Lorentz factor of the blast wave (which changes with time) and of the density of the circumburst medium at the radius of the blast wave (which may or may not change with time). This introduces a break

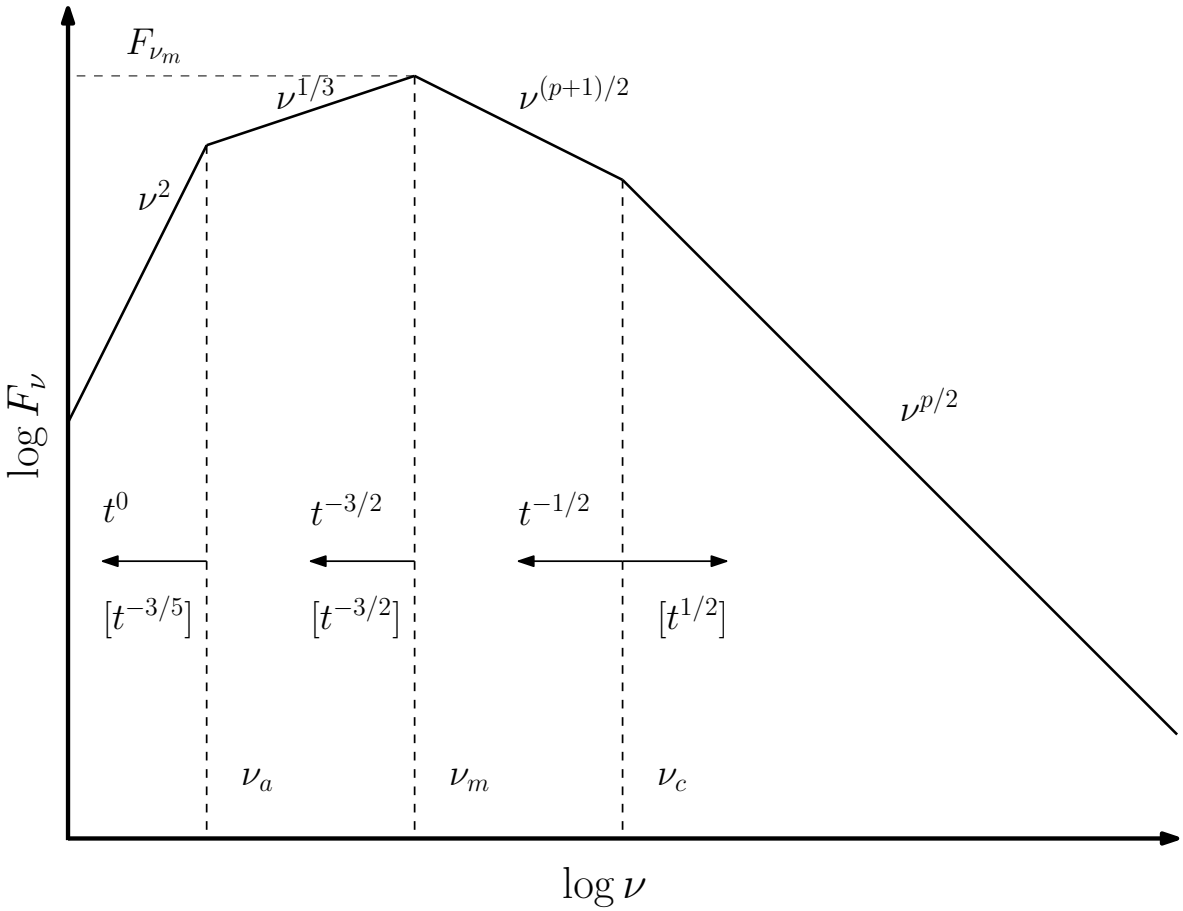

Figure 3.1 Theoretical synchrotron spectrum of a GRB afterglow for the standard blast wave model for $\nu_a < \nu_m < \nu_c$, where ν_a is the synchrotron self-absorption break, ν_m is the synchrotron peak, and ν_c is the cooling break frequency. These spectral breaks, and the peak flux density F_{ν_m}, are indicated with dashed lines. Arrows indicate the time evolution of the spectral breaks; scalings above the arrows are those for a constant-density circumburst medium ($k = 0$), and those below the arrows are for a wind-driven ($k = -2$) medium. Spectral indices between the breaks are indicated; for $\nu > \nu_m$, the indices are functions of the electron energy index $p < -2$. Smoothing of the breaks is not shown here.

in the spectrum, called the "cooling break" at frequencies $\nu > \nu_c(t)$, where $\nu_c(t)$ is the characteristic synchrotron frequency of an electron with energy $E_c(t)$. At frequencies $\nu > \nu_c$, the spectrum varies as $\nu^{p/2}$ (e.g., Sari et al., 1998). For a blast wave expanding into a constant-density medium, the cooling break frequency decreases with time as $t^{-1/2}$. For a wind-driven circumburst density profile $n(r) \propto r^{-2}$, ν_c *increases* with time as $t^{1/2}$ (e.g., Chevalier & Li, 1999); as the shock propagates outward through a medium of ever-decreasing density, the high-energy tail of the electron energy distribution is actually replenished. In general, for a circumburst medium with a power-law radial density profile $n(r) \propto r^k$, the time index of the cooling break frequency is a function of k (e.g., Starling et al., 2008). In typical GRB afterglows, the cooling break frequency lies above the optical/NIR region of the spectrum at early times. If the circumburst medium is, e.g., of constant or near-constant density ($k \simeq 0$), the cooling break may then pass through the optical/NIR before the afterglow fades completely from view; hence, the cooling break is an important part of our afterglow emission model.

Lastly, at low frequencies $\nu < \nu_a$, synchrotron self-absorption causes another break in the spectrum, below which the spectrum varies as ν^2. Note that this differs from the $\nu^{5/2}$ spectral index derived by Rybicki & Lightman (1979), since the low-energy spectrum is dominated primarily by emission from electrons with energy $E \sim E_m$ (Katz, 1994). The self-absorption break frequency ν_a does not change with time in the standard adiabatic blast wave model in a constant density circumburst medium (e.g., Sari et al., 1998), and decreases as $t^{-3/5}$ for a wind-blown medium (e.g., Chevalier & Li, 1999). In typical GRB afterglows, ν_a lies in the radio region of the spectrum, at frequencies of a few GHz.

The transitions between power law spectral regimes across the various breaks are predicted to be smooth. If a break occurs at frequency ν_B, and the spectral indices at

frequencies $\nu < \nu_B$ and $\nu > \nu_B$ are β_1 and β_2, respectively, then:

$$F_\nu = F_{\nu_B} \left[\left(\frac{\nu}{\nu_B} \right)^{-s_B \beta_1} + \left(\frac{\nu}{\nu_B} \right)^{-s_B \beta_2} \right]^{-1/s_B} , \tag{3.2}$$

where F_{ν_B} is the (unsmoothed) flux density at the frequency of the break, and where $s_B(p, k)$ is a smoothing parameter that is a function of the electron energy index p and circumburst density index k, depending on the break in question (e.g., Granot & Sari, 2002). Smoothing parameter values $s_B \gg 1$ correspond to very sharp breaks, while values $0 < s_B \ll 1$ correspond to very gradual breaks in the spectrum.

To describe the spectral flux density of a burst, it is necessary to specify the flux density at a particular frequency, in addition to the break frequencies, power law indices and smoothing parameters. While in practice, this normalizing flux density can be referenced to any frequency, in the literature it is common to use F_{ν_m}, the flux density at the spectral peak frequency ν_m. This normalizing flux F_{ν_m} changes with time as $t^{f(k)}$, where $f(k)$ is a function of the circumburst density index k. In the cases of adiabatic expansion in constant density and wind-blown circumburst density profiles, $f(0) = 0$ and $f(-2) = -3/2$, respectively (e.g., Mészáros & Rees, 1997; Sari et al., 1998; Chevalier & Li, 1999). Once F_{ν_m} is known, the flux at the other breaks F_{ν_a} and F_{ν_c} can be computed from it, using the known spectral indices between the breaks.

From the measured spectral and temporal indices of a GRB afterglow, we can compute the electron energy index p and the circumburst radial density profile index k. If the spectral and temporal coverage of our observations allows us to measure the frequencies of the three spectral breaks ν_m, ν_c and ν_a, and the normalizing flux at one of these breaks, e.g., F_{ν_m}, we can compute additional physical parameters: the isotropic-equivalent energy of the blast wave, $E_{\rm iso}$; the fraction of that energy that resides in electrons and in the magnetic field, ϵ_e and ϵ_B; and the circumburst density

normalization, e.g., n or A_* for $k = 0$ or $k = -2$, respectively (e.g., Sari et al., 1998; Chevalier & Li, 1999). If any of the three break frequencies or the flux normalization are not known, it is still possible, by modeling a discontinuity in the afterglow light curve or spectrum (e.g., a rebrightening) to compute the proportional changes in these physical quantities that would be necessary to explain the observed discontinuity, and possibly rule one or more of them out (see §4.3.1).

3.2.2 Implementation of the Standard Blast Wave Model

At observing frequencies that lie in between any two spectral breaks, the emitted flux density as a function of frequency and time can be expressed in terms of simple power law relations:

$$F_\nu(t) = F_0 \left(\frac{t}{t_0}\right)^\alpha \left(\frac{\nu}{\nu_0}\right)^\beta , \tag{3.3}$$

where F_0 is the model flux density at some (arbitrary) reference time t_0, at some reference frequency ν_0 that lies between the two breaks at time t_0. Depending on which spectral break frequencies that ν is between, the spectral index β may be a function of the electron energy index p, and the temporal index α may be a function of both p and the circumburst density index k (e.g., Starling et al., 2008).

There are two special regimes of the afterglow light curve that we must treat differently from the standard decelerating blast wave model described in §3.2.1. The first is at early times when the afterglow is "turning on", i.e., when the light curve is rising rather than falling. At very early times, there is expected to be no afterglow emission, until the deceleration time, at which the expanding jet has swept up a mass of circumburst material $M \sim M_{\rm ej}/\Gamma$, where $M_{\rm ej}$ is the mass of the ejecta and Γ is the bulk Lorentz factor of the jet. Before this time, the ejecta are effectively coasting outwards unimpeded. After this time, the jet begins to decelerate, depositing energy into the circumburst medium in the form of an ultrarelativistic forward shock (and sometimes

also a reverse shock; e.g., Sari et al. (1999)), which evolves and emits radiation with a fading light curve, according to the standard blast wave model. But the physics of the emission during the transition, when the light curve is, by necessity, rising, is poorly understood theoretically. Modeling this "ramp-up" phase is further complicated by the fact that there are a number of other processes that can result in a rising light curve, including: our relativistic viewing cone broadening to overlap a potentially brighter central structure of an off-axis jet as it decelerates; the blast wave moving through a region of increasing density; and energy injection by a "refreshed shock" due to later-time ejecta catching up with the decelerating blast wave (e.g., Perley et al., 2009). In fact, the latter two processes can result in rebrightening episodes well after the afterglow has entered the standard decelerating blast wave phase. We therefore model the rising light curve in this regime (and in any later rebrightening events) purely empirically. If the temporal index of the post-peak, fading light curve is $\alpha_2 < 0$, then the temporal index during the pre-peak, rising phase is $\alpha_1 = \alpha_2 + \Delta\alpha$, where $\Delta\alpha > -\alpha_2$. We assume that all spectral indices are the same before and after the peak in the light curve, i.e., that $\beta_1 = \beta_2$. Under these assumptions, any spectral break whose frequency scales as $\nu_B(t)$ in the post-peak regime will also scale as $\nu_B(t)$ in the pre-peak regime.

The second special regime of the light curve is at late times when the afterglow is "turning off". After the expanding jet has swept up a mass of circumburst material $M \sim M_{\rm ej}$, its forward motion is strongly checked, and it begins to expand laterally. This transition is known as the "jet break", and the behavior of the afterglow spectrum and light curve at times after the jet break are well understood (e.g., Sari et al., 1999). After the jet break, the temporal index of the light curve is equal to the electron energy index, $\alpha_3 = p$. The frequencies of the various breaks in the spectrum cease to change with time, and all the spectral indices are the same before and after the jet break, i.e., that $\beta_3 = \beta_2$.

We are now in a position to construct a full model of the intrinsic flux density of the afterglow as a function of frequency and time. For simplicity, we will first assume that there is only one spectral break in the data, which occurs at a frequency $\nu_B(t)$. We begin by computing the model flux density in each of the three regimes: 1) the "turn on" phase; 2) the standard blast wave phase; and 3) the jet break phase, independently of one another, as if the light curve in each described the emission at all times. In regime i, we denote the temporal indices at frequencies below and above the break frequency as α_{a_i} and α_{b_i}, respectively, and the spectral indices below and above the break frequency as β_{a_i} and β_{b_i}, respectively. Given the flux F_0 at the reference time t_0 and reference frequency ν_0, and the spectral and temporal indices, we can compute the (unsmoothed) flux at the break, $F_{B_i}(t)$. (Care must be taken to use the correct indices, depending on which side of the break ν_0 lies at times t_0 and t.) Then the model light curve in regime i at observing frequency ν is given by:

$$F_i(t) = F_{B_i}(t) \left[\left(\frac{\nu}{\nu_B(t)} \right)^{-\beta_{a_i} s_B} + \left(\frac{\nu}{\nu_B(t)} \right)^{-\beta_{b_i} s_B} \right]^{-1/s_B} , \tag{3.4}$$

where the smoothing parameter s_B is a function of p and k, depending on which break we are modeling. Equation 3.4 gives us the independent model light curves $F_1(t)$, $F_2(t)$ and $F_3(t)$ in each of the three regimes.

We then combine these independent model light curves, assuming the actual light curve is smoothly broken. If our data only span two regimes, say 1 and 2, then the combined light curve is:

$$F_{12}(t) = \left[F_1(t)^{-s_{12}} + F_2(t)^{-s_{12}} \right]^{-1/s_{12}} , \tag{3.5}$$

where s_{12} is a purely empirical smoothing parameter that must be fit to in our modeling.

Similarly, if our data only span regimes 2 and 3, then the combined light curve is:

$$F_{23}(t) = \left[F_2(t)^{-s_{23}} + F_3(t)^{-s_{23}}\right]^{-1/s_{23}} , \tag{3.6}$$

where, again, s_{23} is a purely empirical smoothing parameter that must be fit to.

If our data span all three regimes, then there are two different ways we can combine the light curves:

$$F_{1,23}(t) = \left[F_1(t)^{-s_{12}} + F_{23}(t)^{-s_{12}}\right]^{-1/s_{12}} , \tag{3.7}$$

or:

$$F_{12,3}(t) = \left[F_{12}(t)^{-s_{23}} + F_3(t)^{-s_{23}}\right]^{-1/s_{23}} . \tag{3.8}$$

If t_{12} and t_{23} are sufficiently well-separated, or if s_{12} and s_{23} are sufficiently > 1, these two combinations yield very similar light curves. But, in general, we compromise and average the two, in $\log F_\nu$ space, to obtain our final model light curve:

$$\log F_{123} = \frac{1}{2}\left(\log F_{12,3} + \log F_{1,23}\right) . \tag{3.9}$$

Finally, we note that we could, of course, have a light curve with multiple spectral breaks. In that case, the model flux density $F_i(t)$ of Equation 3.4 has to be generalized, along the same lines as Equation 3.9.

3.3 Dust Extinction Models

We now present our models for extinction due to dust in both the Milky Way and in the source frame of the GRB. For Milky Way extinction, we employ the empirical model of Cardelli et al. (1989, hereafter CCM; §3.3.1), which is effectively a function of a single free model parameter, $R_V^{\rm MW}$. In §3.3.2, we review previously published R_V

measurements along Milky Way lines of sight, and construct a parameterized model for the probability distribution $p(\log R_V^{\rm MW})$. We then fit this model to the data to obtain probability distributions on those model parameters, which in turn we employ as prior constraints on our Milky Way extinction model.

For dust extinction in the source frame of the GRB, we employ a combination of the empirical CCM model and that of Fitzpatrick & Massa (1988, hereafter FM; §3.3.3). In §3.3.4, we review previously published published measurements of FM extinction parameters along Milky Way and Magellanic Cloud lines of sight, and construct parameterized models describing the probability distributions (and, in some cases, correlations) of these parameters. We then fit these models to the data to obtain probability distributions on those model parameters, which in turn we employ as prior constraints on our source-frame extinction model.

3.3.1 Milky Way Dust Extinction Model

The empirical CCM dust extinction model describes the extinction, in magnitudes, at a given wavelength:

$$A_\lambda^{\rm CCM} = A_V \left[a(x) + \frac{b(x)}{R_V} \right] . \tag{3.10}$$

Here, A_V is the extinction, in magnitudes, in the V band. The parameter $R_V \equiv \frac{A_V}{E(B-V)}$, where the color excess $E(B-V) \equiv A_B - A_V$. The wave number $x \equiv (\lambda/1\ \mu{\rm m})^{-1}$, and $a(x)$ and $b(x)$ are empirical functions that vary depending on the wavelength range in question. For reference, radiation frequency $\nu/(10^{14}\,{\rm Hz}) \simeq 3x$. For $(x < 1.1)$:

$$\begin{aligned} a(x) &= 0.574x^{1.61} \\ b(x) &= -0.527x^{1.61} . \end{aligned} \tag{3.11}$$

For $(1.1 < x < 3.3)$:

$$\begin{aligned} a(x) &= 1 + 0.17699y - 0.50447y^2 - 0.02427y^3 + 0.72085y^4 \\ &\quad +0.01979y^5 - 0.77530y^6 + 0.32999y^7 \\ b(x) &= 1.41338y + 2.28305y^2 + 1.07233y^3 - 5.38434y^4 \\ &\quad -0.62251y^5 + 5.302060y^6 - 2.09002y^7\,, \end{aligned} \tag{3.12}$$

where $y \equiv x - 1.82$. For $(3.3 < x < 8.0)$:

$$\begin{aligned} a(x) &= 1.752 - 0.316x - 0.104/\left[(x-4.67)^2 + 0.341\right] + A(x) \\ b(x) &= -3.090 + 1.825x + 1.206/\left[(x-4.62)^2 + 0.263\right] + B(x)\,, \end{aligned} \tag{3.13}$$

where

$$\begin{aligned} A(x) &= \begin{cases} 0 & \text{if}\, x < 5.9 \\ -0.04473(x-5.9)^2 - 0.009770(x-5.9)^3 & \text{if } x > 5.9 \end{cases} \\ B(x) &= \begin{cases} 0 & \text{if}\, x < 5.9 \\ 0.2130(x-5.9)^2 + 0.1207(x-5.9)^3 & \text{if } x > 5.9 \end{cases}. \end{aligned} \tag{3.14}$$

Finally, for $(8.0 < x < 10.97)$:

$$\begin{aligned} a(x) &= -1.073 - 0.628(x-8.0) + 0.137(x-8.0)^2 - 0.070(x-8.0)^3 \\ b(x) &= 13.670 + 4.257(x-8.0) - 0.420(x-8.0)^2 + 0.374(x-8.0)^2\,. \end{aligned} \tag{3.15}$$

We do not model dust extinction at frequencies above the Lyman limit ($x > 10.97$), since we assume total absorption of the light by neutral hydrogen in that region of the spectrum (see §3.4).

3.3.2 Milky Way Extinction Parameter Priors

In practice, we model Milky Way dust extinction at a given frequency with the parameters $R_V^{\rm MW}$ and $E(B-V)^{\rm MW}$. Since $A_V \equiv R_V E(B-V)$, the extinction along the line of sight due to dust in the Milky Way is:

$$A_\lambda^{\rm MW} = R_V^{\rm MW} E(B-V)^{\rm MW} \left[a(x) + \frac{b(x)}{R_V^{\rm MW}} \right] . \tag{3.16}$$

In $\log F_\nu$ space, the corresponding flux deficit $\Delta \log F_\nu^{\rm MW} = -A_\lambda^{\rm MW}/2.5$. The color excess $E(B-V)^{\rm MW}$ is obtained from interpolation of the all-sky NIR dust emission maps of Schlegel et al. (1998) for the line of sight to a given GRB, and is held as a fixed model parameter. Dust extinction in the Milky Way is thus effectively a function of only one free model parameter, $R_V^{\rm MW}$. However, we do not allow $R_V^{\rm MW}$ to be entirely free. Instead, we have analyzed previous measurements of extinction along various Milky Way lines of sight (see Figure 3.2) to construct a parameterized model for $R_V^{\rm MW}$, with prior constraints on that model's parameters.

Our data set consists of $R_V^{\rm MW}$ measurements towards 417 stars in the Milky Way (Valencic et al., 2004).[2] Since the distribution of $R_V^{\rm MW}$ has a long positive tail, we chose to model it in $\log R_V^{\rm MW}$ space, though we still allow the distribution to be asymmetric in this space.

Since the $R_V^{\rm MW}$ measurements were reported with symmetric, presumably Gaussian, error bars, we transformed their symmetric Gaussian pdfs into $\log R_V^{\rm MW}$ space using the technique described in §2.7, and approximated the transformed intrinsic pdf of data point n as an asymmetric Gaussian with peak $\log R_{V,n}^{\rm MW}$ given by Equation 2.70 and with $\pm 1\sigma$ widths $\sigma_{\log_{R_{V,n}^{\rm MW}} \pm}$ obtained via the bar-lowering technique of §2.6.5 on the

[2]Gordon et al. (2003) report R_V measurements for an additional 23 stars along Magellanic Cloud lines of sight; while we include these stars in our fits to x_0 and γ, and to the correlative c_1 vs. c_2, BH vs. c_2 and R_V vs. c_2 priors for source-frame extinction (see §3.3.4), we omit them in this analysis, which is only intended to describe local extinction along typical Milky Way lines of sight.

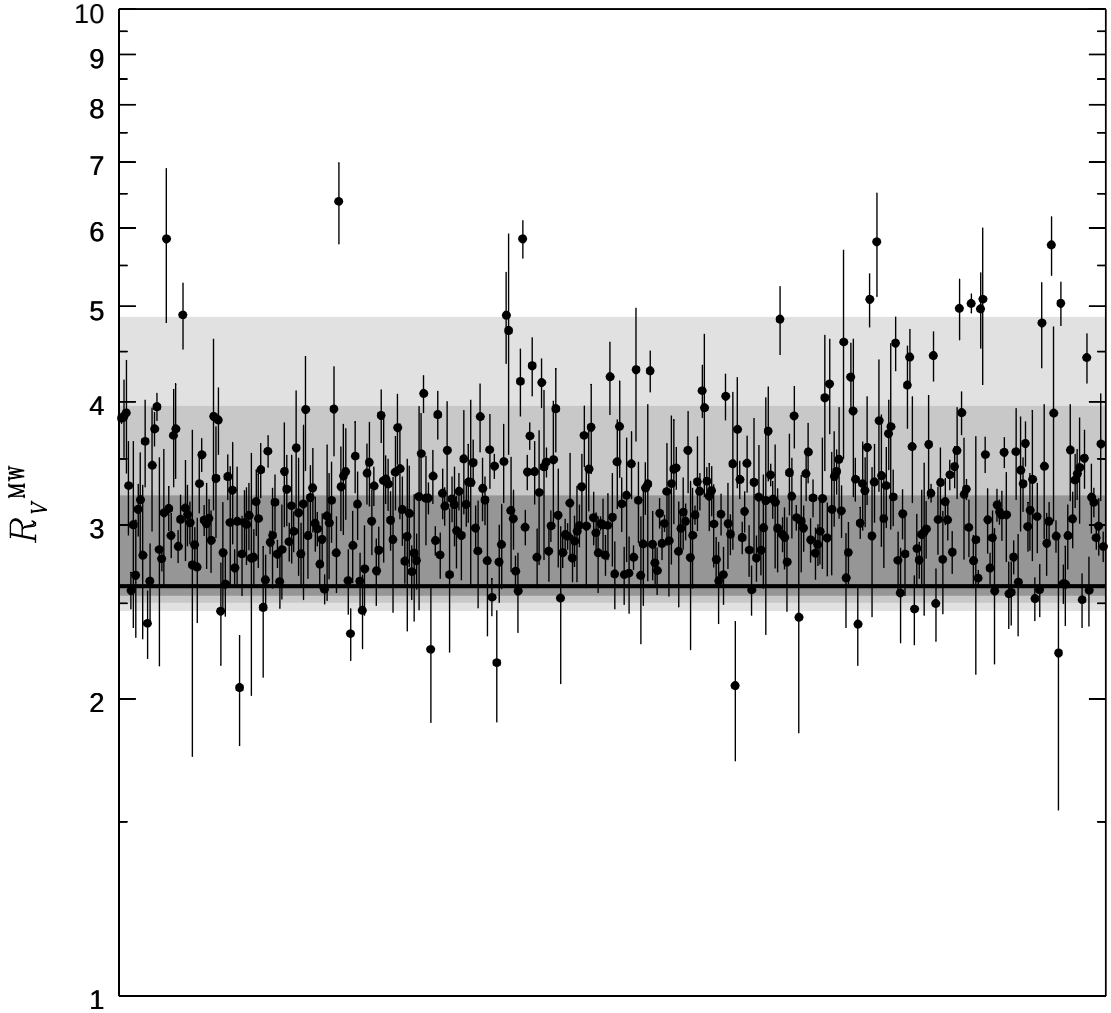

Figure 3.2 Observed $R_V^{\rm MW}$ data, transformed to $\log R_V^{\rm MW}$ space, for 417 Milky Way lines of sight (Valencic et al., 2004), and fitted asymmetric Gaussian model distribution. Data points are plotted versus a randomized index. Shaded regions indicate the 1-, 2- and 3σ envelopes of the model distribution, using the peak prior values of Table 3.1.

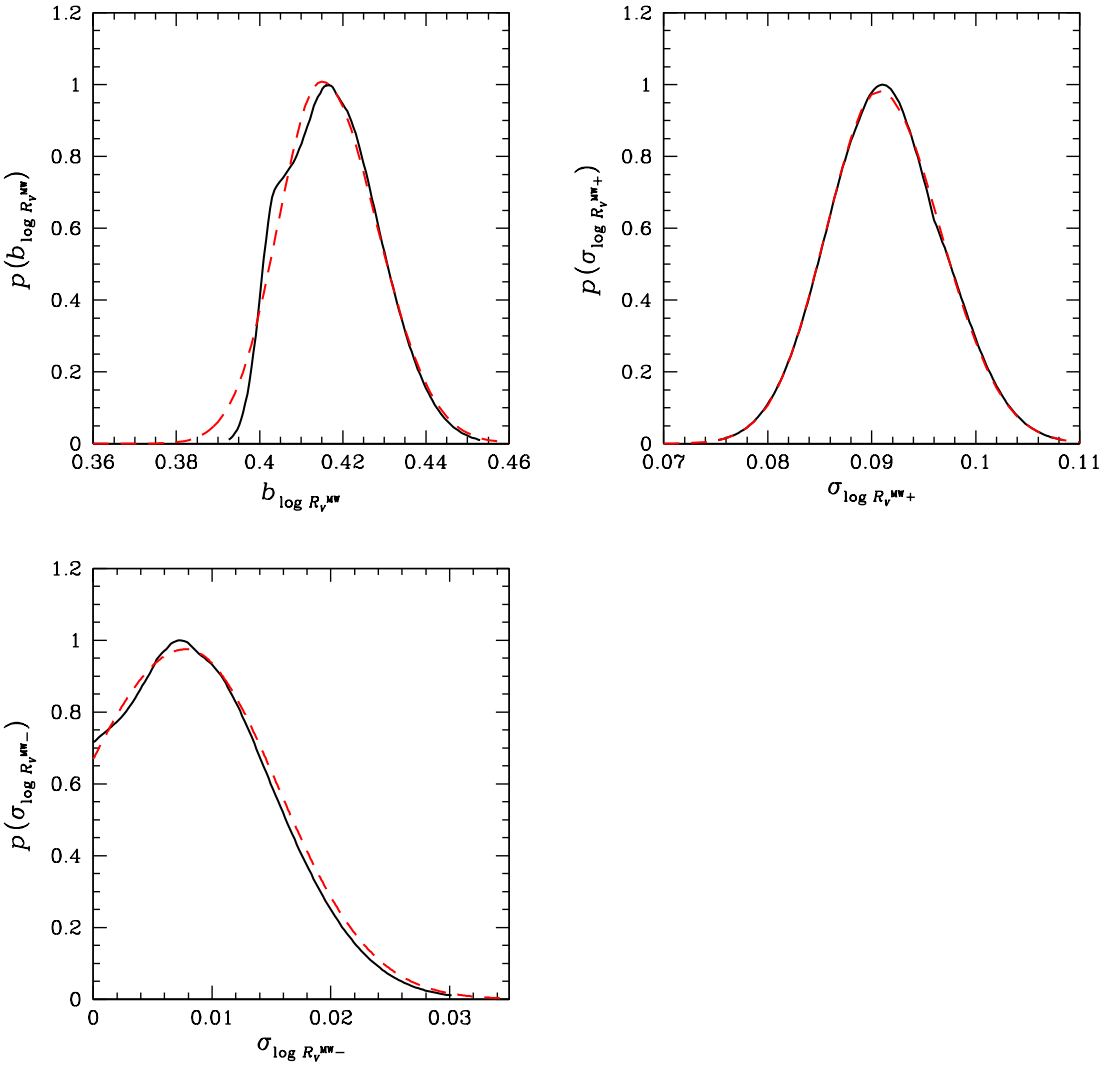

Figure 3.3 Fitted prior probability distributions of $R_V^{\rm MW}$ model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.1

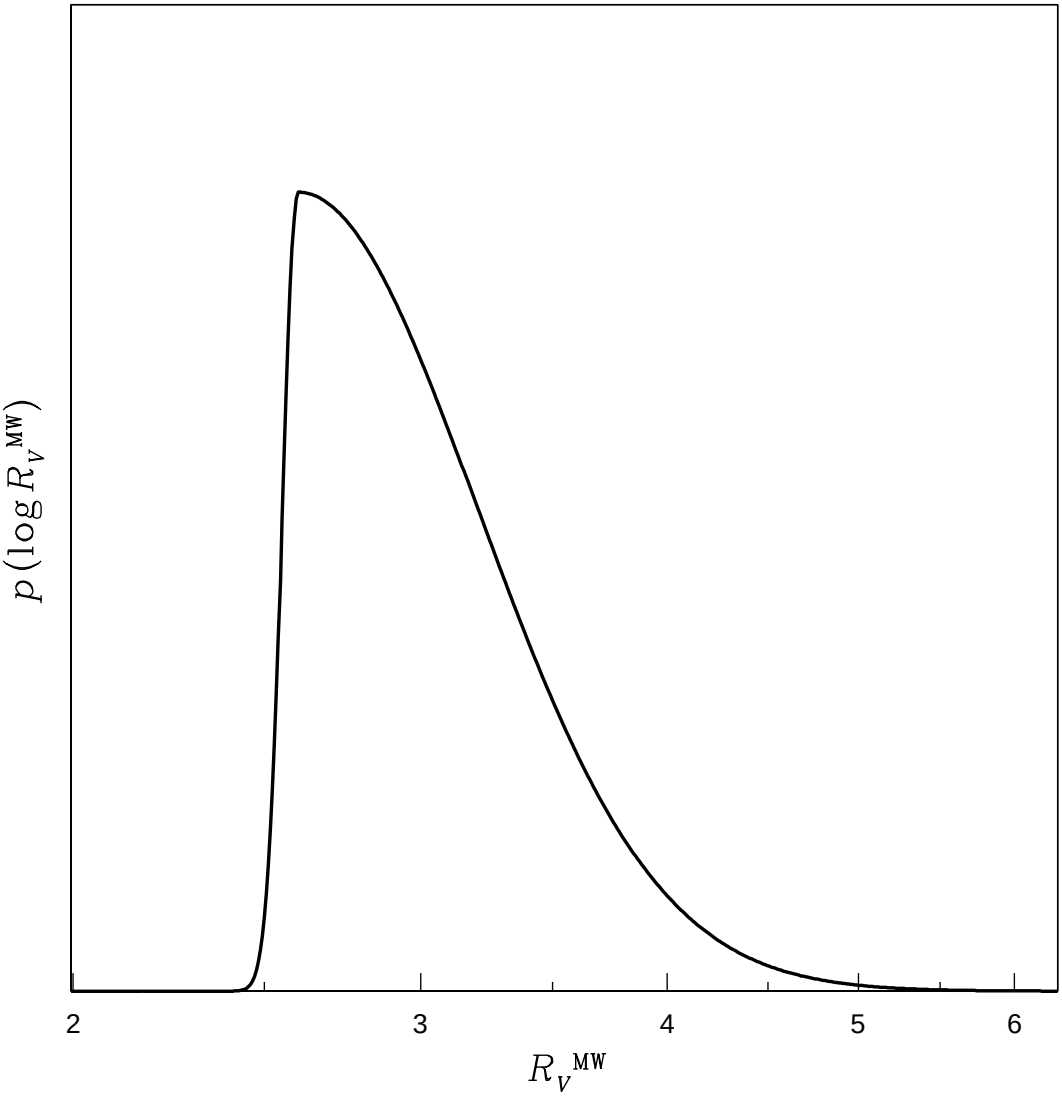

Figure 3.4 Asymmetric Gaussian model distribution fit to the observed distribution of the Milky Way extinction parameter $\log R_V^{\mathrm{MW}}$, using the peak prior parameter values of Table 3.1.

transformed pdf, yielding:

$$p_n^{\rm int} = G_A(\log R_V^{\rm MW}; \log R_{V,n}^{\rm MW}, \sigma_{\log^{\rm MW}_{R_{V,n}}\pm}) \tag{3.17}$$

where G_A is an asymmetric Gaussian function (see Equation 2.62 or A.8). Figure 3.2 shows these transformed $\log R_{V,n}^{\rm MW}$ data points, plotted versus a randomized index.

We fit a model distribution to these data of the form $\log R_V^{\rm MW} = \text{constant} = b_{\log R_V^{\rm MW}}$, convolved with asymmetric Gaussian slop with widths $\sigma_{\log R_V^{\rm MW}\pm}$, i.e.:

$$p^{\rm mod} = G_A\left(\log R_V^{\rm MW}; b_{\log R_V^{\rm MW}}, \sigma_{\log R_V^{\rm MW}\pm}\right) . \tag{3.18}$$

Since this is a 1D model, with slop entirely in the $\log R_V^{\rm MW}$ dimension, and since the intrinsic error bars in the data point indices equal zero, the 2D TRF statistic reduces to a 1D χ^2-like statistic (§2.4.2). We approximate the convolution of the asymmetric Gaussian model distribution with the asymmetric Gaussian intrinsic data point pdf as an asymmetric Gaussian, so the resulting joint probability has the form:

$$p_n = G_A\left(\log R_V^{\rm MW}; \log R_{V,n}^{\rm MW} + \delta_n, \Sigma_{\log R_{V,n}^{\rm MW}\pm}\right) , \tag{3.19}$$

where the slop-convolved uncertainties are:

$$\Sigma^2_{\log R_{V,n}^{\rm MW}\pm} \equiv \sigma^2_{\log R_V^{\rm MW}\mp} + \sigma^2_{\log R_{V,n}^{\rm MW}\pm} , \tag{3.20}$$

and where the shift δ_n in the convolved distribution's peak is given by the algorithm described in Appendix A. The best-fit model is obtained by maximizing the likelihood

function:

$$\begin{aligned}\mathcal{L} \;\propto\; & \prod_{n=1}^{N} \frac{2}{\sqrt{2\pi}(\Sigma_{\log R_{V,n}^{\rm MW}+} + \Sigma_{\log R_{V,n}^{\rm MW}-})} \\ \times\; & \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{\log R_{V,n}^{\rm MW}+\delta_n-b_{\log R_V^{\rm MW}}}{\Sigma_{\log R_{V,n}^{\rm MW}+}}\right)^2\right] & \text{if } \log R_{V,n}^{\rm MW}+\delta_n \geq b_{\log R_V^{\rm MW}}, \\ \exp\left[-\frac{1}{2}\left(\frac{\log R_{V,n}^{\rm MW}+\delta_n-b_{\log R_V^{\rm MW}}}{\Sigma_{\log R_{V,n}^{\rm MW}-}}\right)^2\right] & \text{if } \log R_{V,n}^{\rm MW}+\delta_n < b_{\log R_V^{\rm MW}}. \end{cases}\end{aligned} \tag{3.21}$$

We used our genetic algorithm software *Galapagos* (Foster & Reichart, in preparation; see §2.6.3) to obtain best-fit projected probability distributions (§2.6.4) to the model parameters $b_{\log R_V^{\rm MW}}$, $\sigma_{\log R_V^{\rm MW}+}$ and $\sigma_{\log R_V^{\rm MW}-}$. Figure 3.3 shows the resulting probability distributions. Ideally, we would employ these distributions directly to define the priors on each parameter; however, in practice, they would have to be implemented via a lookup table, which is computationally expensive. Instead, we fit asymmetric Gaussian functions to these distributions to obtain approximate analytic priors for each parameter, which are plotted as dashed red curves in Figure 3.3. The peaks and $\pm1\sigma$ widths of these fitted asymmetric Gaussian priors are presented in Table 3.1. The model distributions plotted in Figures 3.2 & 3.4 were computed using these peak values.

We emphasize that the distribution plotted in Figures 3.2 & 3.4 is *not* the prior on $\log R_V^{\rm MW}$, but merely the most probable parameterization of it. Instead, the actual priors are on the secondary parameters $b_{\log R_V^{\rm MW}}$, $\sigma_{\log R_V^{\rm MW}+}$ and $\sigma_{\log R_V^{\rm MW}-}$, and are approximated by asymmetric Gaussian functions fitted to the probability distributions of Figure 3.3:

$$\begin{aligned} p(b_{\log R_V^{\rm MW}}) &= G_A(b_{\log R_V^{\rm MW}}; 0.4150, +0.0132, -0.0105)\,, \\ p(\sigma_{\log R_V^{\rm MW}+}) &= G_A(\sigma_{\log R_V^{\rm MW}+}; 0.09074, +0.00586, -0.00510)\,, \\ p(\sigma_{\log R_V^{\rm MW}-}) &= G_A(\sigma_{\log R_V^{\rm MW}-}; 0.00779, +0.00736, -0.00897)\,. \end{aligned} \tag{3.22}$$

Table 3.1. Best-Fit Asymmetric Gaussian Priors for $\log R_V^{\rm MW}$ Probability Distribution

Parameter	Peak	$+1\sigma$ Width	-1σ Width
$b_{\log R_V^{\rm MW}}$	0.4150	+0.0132	−0.0105
$\sigma_{\log R_V^{\rm MW}+}$	0.09074	+0.00586	−0.00510
$\sigma_{\log R_V^{\rm MW}-}$	0.00779	+0.00736	−0.00897

These three parameters, constrained by their respective priors, are what we fit to in our GRB afterglow modeling. (Note that, in addition to these prior constraints, both slop parameters are constrained to always be $\geq$0. This is particularly important for the slop parameter $\sigma_{\log R_V^{\rm MW}-}$, whose peak is $< 1\sigma$ above 0.)

Finally, the actual value of $R_V^{\rm MW}$ that we employ in our Milky Way dust extinction model is:

$$R_V^{\rm MW} = 10^{b_{\log R_V^{\rm MW}} + \delta_{\log R_V^{\rm MW}}} , \tag{3.23}$$

where $\delta_{\log R_V^{\rm MW}}$ is a dummy model parameter, constrained by a zero-mean asymmetric Gaussian prior with $\pm 1\sigma$ widths given by the above prior-constrained slop parameters $\sigma_{\log R_V^{\rm MW}\pm}$, i.e.:

$$p(\delta_{\log R_V^{\rm MW}}) = G_A(\delta_{\log R_V^{\rm MW}}; 0, \sigma_{\log R_V^{\rm MW}\pm}) . \tag{3.24}$$

3.3.3 Source-Frame Dust Extinction Model

We model dust extinction in the GRB source frame with a combination of the CCM model and the empirical UV model of Fitzpatrick & Massa (1988, hereafter FM). For $x < 1.82$, we employ the CCM model (Equation 3.10). For $x > 3.3$, we employ the FM

model:

$$\frac{E(\lambda - V)}{E(B - V)} = c_1 + c_2 x + c_3 D(x; \gamma, x_0) + c_4 F(x)\,. \tag{3.25}$$

The three functional subcomponents of Equation 3.25 are illustrated with dashed lines in Figure 3.5, and are due to different physical characteristics of the population of dust grains responsible for the extinction (see, e.g. Reichart, 2001). While the extinction at optical wavelengths is thought to be due to large, fluffy ($1000 - 2000$ Å) classical van de Hulst grains (van de Hulst, 1957), extinction in the UV – especially in the far UV – is due to smaller grains, possibly including modified classical grains, with sizes ranging from ≈1000 Å to 100 Å or smaller.

The linear component of the FM extinction model, $c_1 + c_2 x$, has a slope c_2 that is largely determined by the distribution of these smaller dust grain sizes in the population. A flat $c_2 = 0$ corresponds to "gray dust", or a population with uniform, relatively large grains that extinguishes all UV wavelengths equally; such extinction is typical of Orion nebula lines of sight, where stellar winds are thought to have swept away the smaller grains, and is also seen in some other cold, dense clouds, where smaller grains may have accreted onto the larger ones. Slopes of $0.6 \lesssim c_2 \lesssim 1$ are typical of the diffuse Milky Way ISM. A steep $c_2 \gtrsim 1$ is typical of highly processed dust, like that of the older star-forming regions in the Magellanic Clouds, where fragmentation by supernova shocks has resulted in an overabundance of smaller grains.

The next term in Equation 3.25:

$$\begin{aligned} c_3 D(x; \gamma, x_0) &= c_3 \frac{x^2}{(x^2 - x_0^2)^2 + x^2\gamma^2} \\ &= \mathrm{BH} \frac{x^2}{\left(x^2 - x_0^2\right)^2 / \gamma^2 + x^2} \end{aligned} \tag{3.26}$$

describes the Drude profile of the 2175 Å feature, or "UV bump", which peaks at x_0, has an approximate FWHM γ, and a bump height $\mathrm{BH} \equiv c_3/\gamma^2$ in this space. The

origin of the UV bump is not fully understood, but there is evidence that it is due to a different population of grains from that responsible for the linear $c_1 + c_2 x$ extinction (e.g. Greenberg & Chlewicki, 1983), perhaps very small, hydrogen-poor graphitic grains (e.g. Hecht, 1986). Both the parameter c_3 and the bump height BH are observed to be correlated with c_2 . However, fitting to BH is preferred over fitting to c_3, since when measuring c_3 and the bump width γ, their uncertainties are likely correlated, but when measuring BH and γ, their uncertainties should be relatively independent. We therefore chose to fit correlative model distributions to BH vs. c_2 (see §3.3.4.3), and a 1D probability distribution to γ (see §3.3.4.4).

The final term in Equation 3.25 is $c_4 F(x)$, where $F(x)$ is an empirical function:

$$F(x) = \begin{cases} 0 & \text{if } (x < 5.9) \\ 0.5392(x-5.9)^2 + 0.05644(x-5.9)^3 & \text{if } (x > 5.9)\,. \end{cases} \tag{3.27}$$

The model parameter c_4 thus describes the magnitude of far UV extinction in excess of the linear $c_1 + c_2 x$ trend. The origin of this feature of UV extinction spectra is even less well understood than that of the UV bump; it may be due to small (< 100 Å) graphitic grains (Draine & Lee, 1984). For the purposes of our GRB modeling, c_4 additionally serves as an empirical "place holder" for extreme dust fragmentation by the GRB itself along the line of sight (e.g. Lee et al., 2001), and, since it is not observed to be correlated with any of the other FM parameters, we do not constrain its value with any sort of prior (other than requiring $c_4 \geq 0$).

Solving Equation 3.25 for A_λ, given that $R_V \equiv \frac{A_V}{E(B-V)}$, gives the extinction, in magnitudes, of the FM model:

$$A_\lambda^{\mathrm{FM}} = A_V \left\{ 1 + \frac{1}{R_V} \left[c_1 + c_2 x + \mathrm{BH} \frac{x^2}{\left(x^2 - x_0^2\right)^2 / \gamma^2 + x^2} + c_4 F(x) \right] \right\} , \tag{3.28}$$

where the parameters A_V, R_V, BH and c_4 are all constrained to be ≥ 0.

For $1.82 \leq x \leq 3.3$, we scale the CCM model $A_\lambda^{\rm CCM}$ (Equation 3.10) linearly so that it matches the FM model $A_\lambda^{\rm FM}$ at $x = 3.3$:

$$A_\lambda = A_\lambda^{\rm CCM} \left[1 + \frac{\left(A_{3.3}^{\rm FM} - A_{3.3}^{\rm CCM}\right)}{A_{3.3}^{\rm CCM}} \frac{(x - 1.82)}{(3.3 - 1.82)} \right] . \tag{3.29}$$

Figure 3.5 illustrates the combined CCM and FM extinction model for a typical Milky Way line of sight. Figure 3.6 shows model spectral flux densities for a hypothetical GRB at redshift $z_{\rm GRB} = 0$ for a variety of FM extinction parameter values.

3.3.4 Source-Frame Extinction Parameter Priors

We analyzed published extinction parameter measurements towards 417 stars in the Milky Way (Valencic et al., 2004, hereafter VCG) and 23 stars in the Large and Small Magellanic Clouds (Gordon et al., 2003, hereafter G03) to obtain priors on our source-frame dust extinction model parameters. Both surveys fit the FM parameterized model to UV stellar spectra to obtain values for the FM extinction parameters. G03 reports values for: $c_{1,n}$, $c_{2,n}$, $c_{3,n}$, $c_{4,n}$, γ_n, $x_{0,n}$ and $R_{V,n}$, with symmetric intrinsic error bars $\sigma_{c_{1,n}}$, $\sigma_{c_{2,n}}$, $\sigma_{c_{3,n}}$, $\sigma_{c_{4,n}}$, $\sigma_{\gamma,n}$, $\sigma_{x_{0,n}}$ and $\sigma_{R_{V,n}}$. VCG reports certain parameter values normalized by R_V: $(c_1/R_V + 1)_n$, $(c_2/R_V)_n$, $(c_3/R_V)_n$ and $(c_4/R_V)_n$, along with γ_n, $x_{0,n}$ and $R_{V,n}$, with symmetric intrinsic error bars $\sigma_{(c_1/R_V+1),n}$, $\sigma_{(c_2/R_V),n}$, $\sigma_{(c_3/R_V),n}$, $\sigma_{(c_4/R_V),n}$, $\sigma_{\gamma,n}$, $\sigma_{x_{0,n}}$ and $\sigma_{R_{V,n}}$. For these VCG data points, we simply computed values for $c_{1,n} = R_{V,n}\left[(c_1/R_V + 1)_n - 1\right]$, $c_{2,n} = R_{V,n}(c_2/R_V)_n$ and $c_{3,n} = R_{V,n}(c_3/R_V)_n$, and assigned these data points symmetric error bars $\sigma_{c_1,n} = R_{V,n}\sigma_{(c_1/R_V+1),n}$, $\sigma_{c_2,n} = R_{V,n}\sigma_{(c_2/R_V),n}$ and $\sigma_{c_3,n} = R_{V,n}\sigma_{(c_3/R_V),n}$.[3] We also computed the UV bump height for each line of

[3]We did not propagate uncertainty in $R_{V,n}$ into these uncertainties. To do so would result in correlated intrinsic uncertainties in the c_1 vs. c_2, R_V vs. c_2 and BH vs. c_2 data sets, which is not a situation the TRF statistic is currently able to address. Instead, we chose to use these *underestimated* intrinsic uncertainties; this results in *overestimated* fitted extrinsic scatter (slop) in our 2D fits, i.e., a

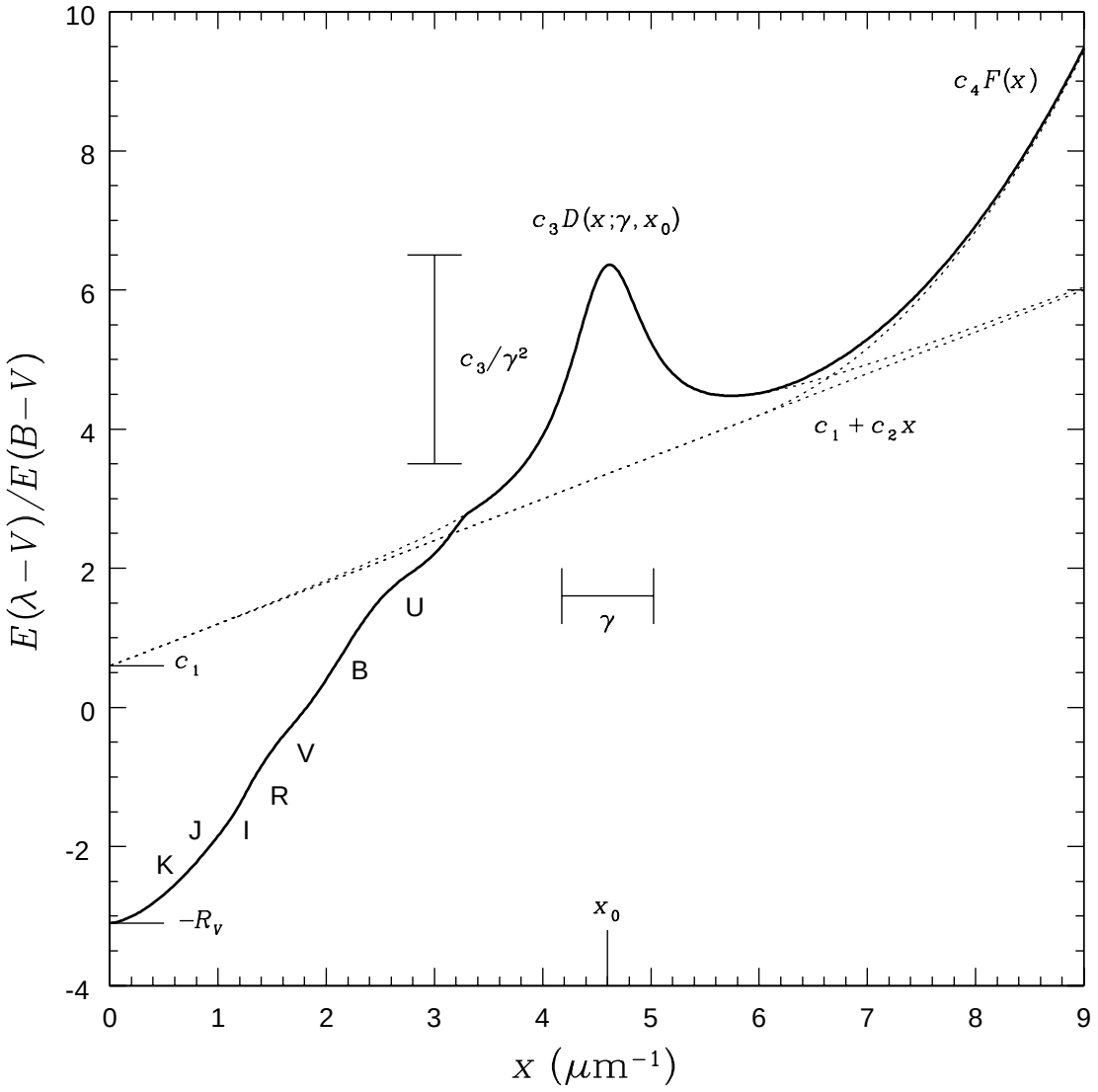

Figure 3.5 Combined CCM and FM dust extinction model, for a typical Milky Way line of sight, with $c_1 = 0.6$, $c_2 = 0.6$, $R_V = 3.1$, $x_0 = 4.6$, $\gamma = 0.85$, BH $\equiv c_3/\gamma^2 = 3$, and $c_4 = 0.5$. For $x \equiv (\lambda/1\ \mu\mathrm{m})^{-1} < 1.82$, extinction is described by the CCM model, and for $x \geq 3.3$ by the FM model. For $1.82 \leq x < 3.3$, A_λ^{CCM} is linearly scaled so as to equal A_λ^{FM} at $x = 3.3$. The various components of the FM extinction model curve are indicated by dotted lines.

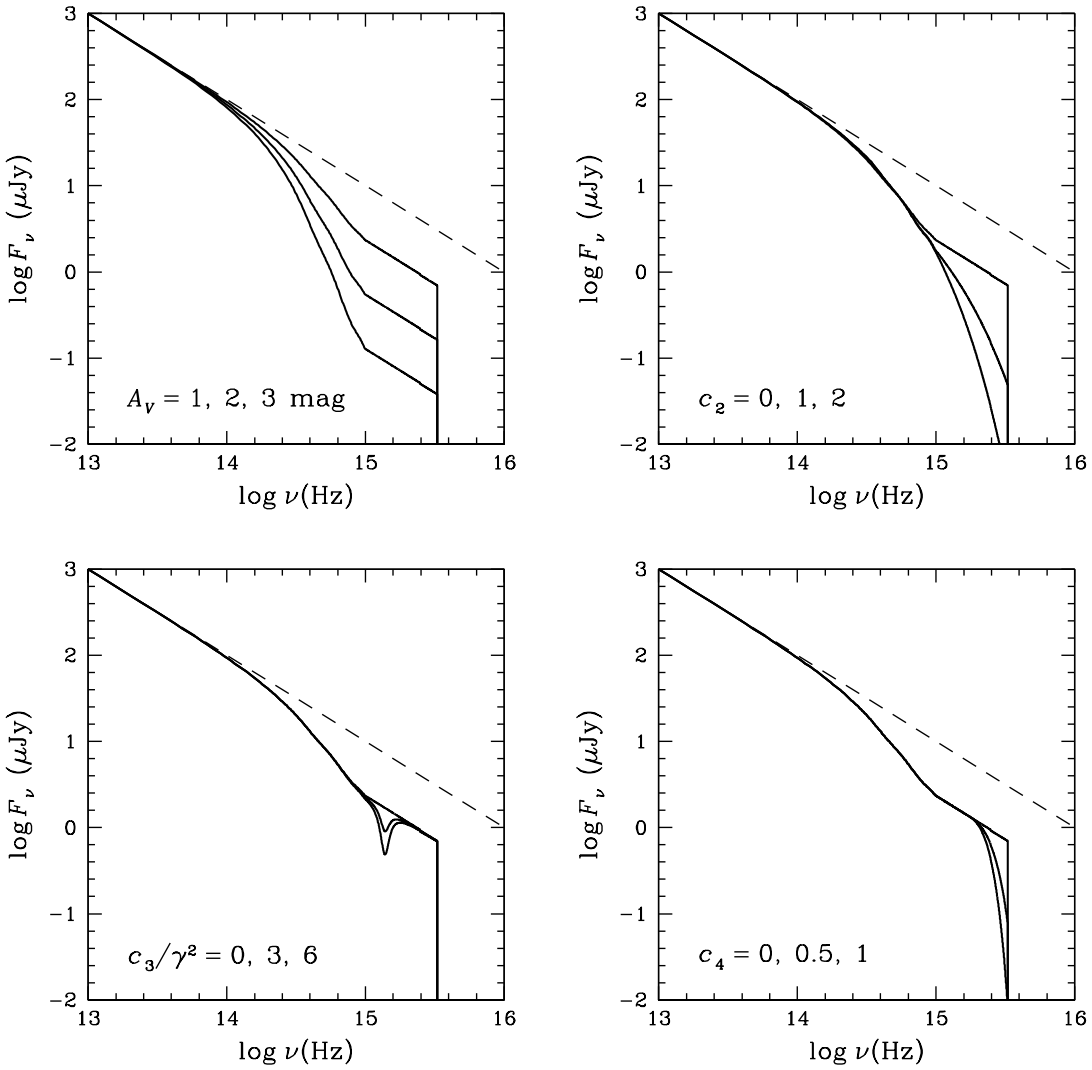

Figure 3.6 Model extinguished spectral flux densities for a GRB at redshift $z_{\rm GRB} = 0$, with an intrinsic power-law spectrum $F_\nu \propto \nu^{-1}$ (dashed line). Default extinction parameter values are $A_V = 1$, $c_2 = 0$, BH $\equiv c_3/\gamma^2 = 0$, $c_4 = 0$. Parameters c_1 and R_V are taken to be functions of c_2, given by the peaks of the best-fit priors described in §3.3.4: For $c_2 = 0$, $R_V = 4.44$ and $c_1 = 2.57$. Uppermost solid curve in each panel shows the model spectrum for the default extinction parameter values. Lower curves show the effect of increasing parameters A_V, c_2, BH, and c_4, respectively. Note that total absorption is assumed at frequencies greater than the Lyman limit, $\log \nu_{\rm LL} = 15.52$.

sight, $\mathrm{BH}_n = c_{3,n}/\gamma_n^2$, and assigned it symmetric error bars using a simple first-order error propagation formula:

$$\sigma_{\mathrm{BH},n} = \left[\left(\frac{2c_{3,n}\sigma_{\gamma,n}}{\gamma_n^3} \right)^2 + \left(\frac{\sigma_{c3,n}}{\gamma_n^2} \right)^2 \right]^{1/2} . \tag{3.30}$$

We treat c_4 as a free model parameter in our GRB model fits, subject only to the constraint $c_4 \geq 0$, and do not analyze the observed $c_{4,n}$ data set further here. For the parameters x_0 and γ, which describe the frequency and width of the UV bump, respectively, we fit simple 1D asymmetric Gaussian model distributions to the measured values (see §3.3.4.4), and obtained fitted priors on those model distributions' parameters, in the same manner as R_V^{MW} (§3.3.2). The parameters c_1, R_V and BH, however, have long been noted to be correlated with the parameter c_2 (e.g. Reichart, 2001; Fitzpatrick & Massa, 1988; Jenniskens & Greenberg, 1993). For these, we fit 2D empirical model distributions to the correlations, using the full TRF statistic. In our source-frame dust extinction model, c_2 is a completely free parameter, while c_1, R_V and BH are functions of c_2 and other parameters, constrained by priors, that describe these fitted correlations (see §§3.3.4.1, 3.3.4.2 & 3.3.4.3).

Before we can proceed to these model distribution fits, however, we must discuss the observational selection function of the combined G03 and VCG data. The selection function is a measure of the efficiency with which a given data set samples the 2D parameter space. Our combined VCG/G03 data set is obviously dominated by typical MW lines of sight, with only a few data points at high c_2 values (typical of star-forming regions whose dust has been heavily fragmented by multiple SNe explosions, as in the Magellanic Clouds), or at low c_2 values (typical of young star-forming regions like the Orion Nebula, where stellar winds have swept smaller dust grains away). Being few in number, these high- and low-c_2 data points are treated as outliers in any 2D model

conservative fitted model distribution.

fits, and the TRF statistic "sees" the data set as something akin to a random cloud of points, at least for the R_V vs. c_2 and BH vs. c_2 fits. For these fits, this results in effectively linear model distributions that fail to reproduce the observed trends in the data for the sparsely sampled high- and low-c_2 regions. But it is precisely in such "atypical" regions that we expect GRBs to preferentially occur. We therefore chose to compute an approximate selection function for our data set, based on the observed distribution of c_2 values. In effect, this amounts to re-weighting each data point. There are many ways one could attempt to quantify the weights. The important point is that *any* reasonable re-weighting scheme is an improvement over uniform weighting.

We chose to assign weights to each data point n that are inversely proportional to the integral of the sum of all N of the Gaussian c_2 intrinsic pdfs, weighted by the $c_{2,n}$ pdf $p_n(c_2) = G(c_2; c_{2,n}; \sigma_{c_2,n})$ of that data point:

$$w_n \propto \left[\int_{-\infty}^{\infty} \left(p_n(c_2) \times \sum_{i=1}^{N} p_i(c_2) \right) dc_2 \right]^{-1} . \tag{3.31}$$

In practice, we normalize the weights so that the minimum weight equals one, which results in weights at the extremes of c_2 of order $\sim$100 (Figure 3.7). We note, again, that we only apply these weights in our computation of priors for source-frame extinction parameters; for the prior on the local extinction parameter $R_V^{\rm MW}$ (§3.3.1), we employed uniform weighting (and omitted the G03 Magellanic Cloud data).

3.3.4.1 Priors on c_1 vs. c_2

The linear component of the FM UV extinction model is described by a slope, c_2, and an intercept c_1 (see Figure 3.5). Through a "lever arm" effect, c_1 is strongly linearly correlated with c_2 in fits to extinction curves for stellar spectra (just as in all linear fits; see §2.6.6). The observed c_1 vs. c_2 data set is plotted in Figure 3.8; the linear

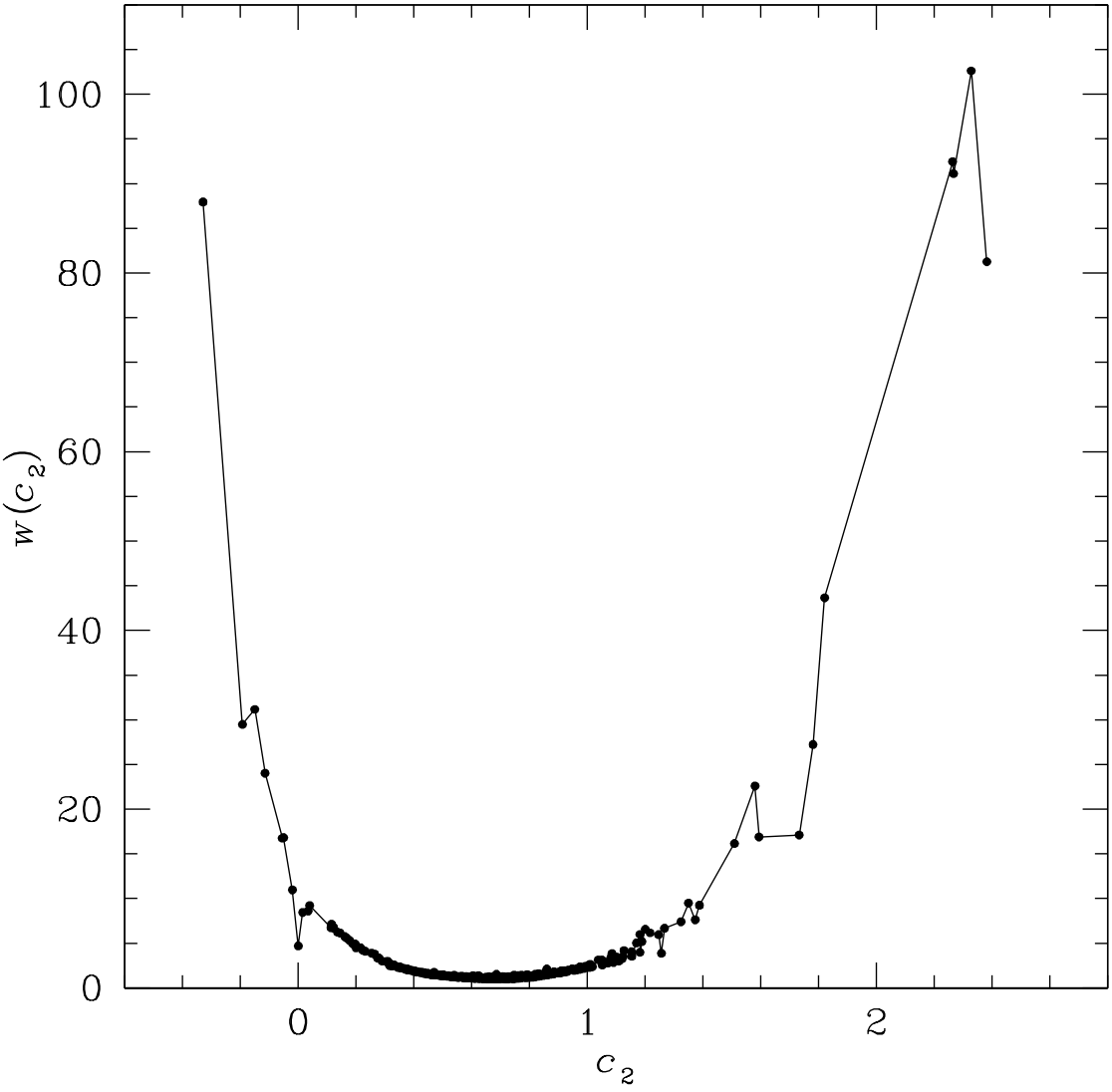

Figure 3.7 Weights as a function of c_2, employed in fits to obtain priors for source-frame dust extinction parameters.

correlation of the data is obvious.

We employed the full 2D TRF statistic to fit linear model distributions to the weighted c_1 vs. c_2 data set, and to find the optimal scale for the fit (§2.5). At any fixed scale, the best-fit extrinsic scatter, or slop, in the x- and y dimensions is described by the parameters $(\sigma_{c_2}^{c_1}, \sigma_{c_1})$, and the model distribution is defined as the convolution of a 2D Gaussian with these widths with a linear model curve:

$$c_{1,c}(c_2; \vartheta_m) = b^{c_1} + \tan\theta^{c_1} \left(c_2 - c_2^{p_{c_1}}\right) . \tag{3.32}$$

Through the iterative technique described in §2.5, we determined the optimum scale for the fit, $s_0 = 0.25767$, at which the correlation coefficient $R^2_{\rm TRF} = 0.985$. From a fine-scale march in θ^{c_1} about the best fit, we determined a value of the pivot point $c_2^{p_{c_1}} = 1.2403$ (§2.6.6). We performed error bar marches (§2.6.5) to determine the projected probability distributions of parameters b^{c_1} and θ^{c_1}. Figure 3.9 shows our fitted $p(b^{c_1})$ and $p(\theta^{c_1})$. To obtain approximate analytic priors on these parameters, we fit asymmetric Gaussian functions to those distributions; these asymmetric Gaussian prior peaks and widths are given in Table 3.2. The model distribution plotted in Figure 3.8 was computed using the peak prior model parameter values of Table 3.2; the 1-, 2- and 3σ envelopes were computed using the technique described in Appendix C. It is important to note, just as before, that these envelopes do *not* represent the actual correlative prior on c_1 vs. c_2, but simply show the most probable fitted model distribution.

In our GRB model, c_2 is fit to as a free parameter, while the parameters b^{c_1} and θ^{c_1} are constrained by the asymmetric Gaussian priors in Table 3.2, and the slops $\sigma_{c_2}^{c_1}$ and σ_{c_1} are held fixed at their best-fit values at the optimum scale s_0.[4] The value of c_1 we

[4]Since the way slop is distributed into the x- and y directions depends entirely on the choice of scale, and since the scale must be held fixed for a given 2D TRF fit, their values are not physically meaningful and, consequently, it is not meaningful to compute projected uncertainties on them.

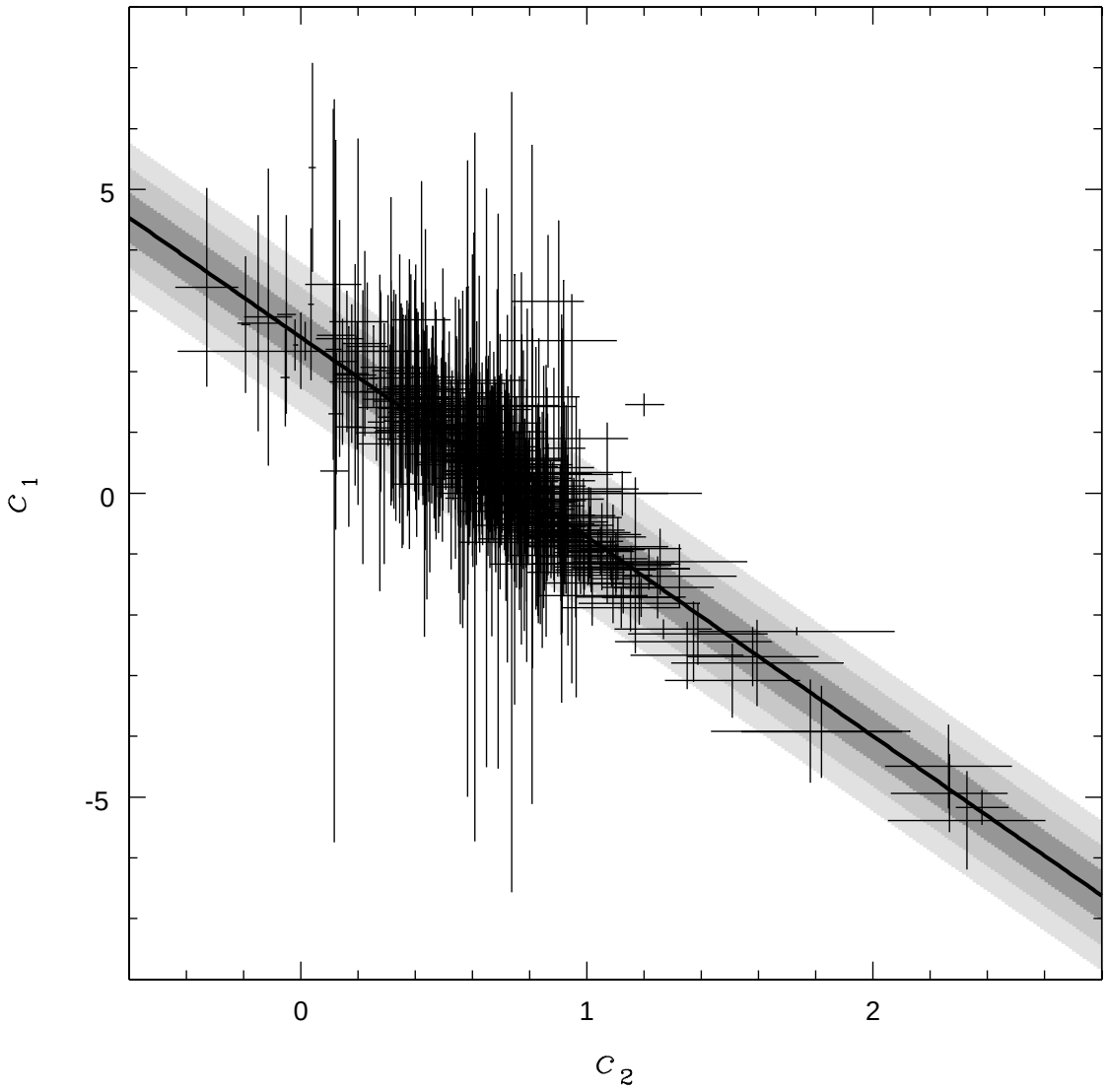

Figure 3.8 Observed c_1 vs. c_2 data from Gordon et al. (2003) and Valencic et al. (2004), and fitted linear model distribution. Shaded regions indicate the 1-, 2- and 3σ slop envelopes of the model distribution, using the peak prior values of Table 3.2, and computed using the technique described in Appendix C.

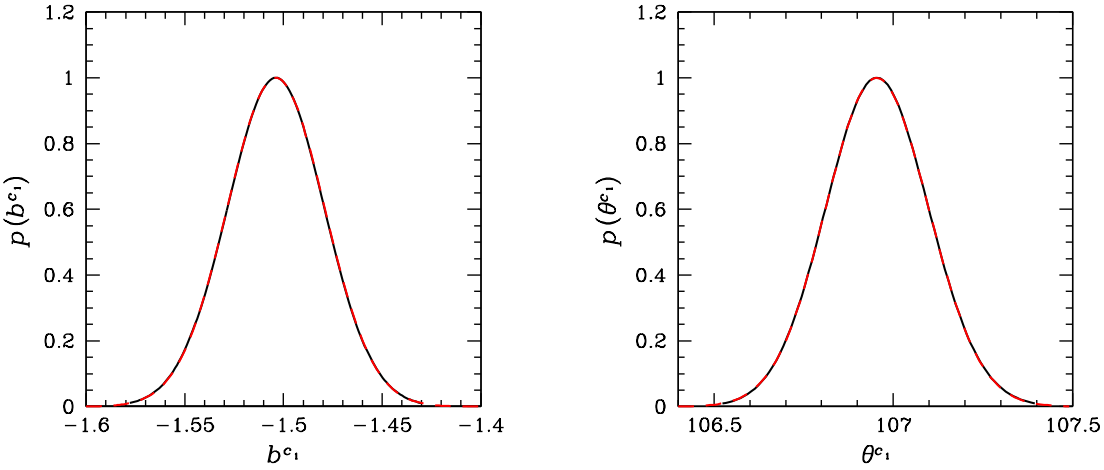

Figure 3.9 Fitted prior probability distributions of c_1 vs. c_2 model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.2

employ in our GRB model fits is:

$$c_1 = c_{1_c}(c_2 - \delta_{c_2}^{c_1}; b^{c_1}, \theta^{c_1}) + \delta_{c_1} \,, \tag{3.33}$$

where the curve $c_{1,c}(c_2; b^{c_1}, \theta^{c_1})$ is given by Equation 3.32, and where $\delta_{c_2}^{c_1}$ and δ_{c_1} are dummy model parameters, constrained by zero-mean Gaussian priors with widths given by $\sigma_{c_2}^{c_1}$ and σ_{c_1}, i.e., $p(\delta_{c_2}^{c_1}) = G(c_2; 0, \sigma_{c_2}^{c_1})$ and $p(\delta_{c_1}) = G(c_1; 0, \sigma_{c_1})$.

3.3.4.2 Priors on R_V vs. c_2

The parameter R_V is observed to be loosely correlated with c_2: Lines of sight with extremely small values of $c_2 \approx 0$, which are typical of young star-forming regions like the Orion Nebula, often show values of $R_V \gtrsim 5$, while R_V settles closer to its canonical value of 3.1 for $c_2 \gtrsim 0.5$. The data are plotted in Figure 3.10.

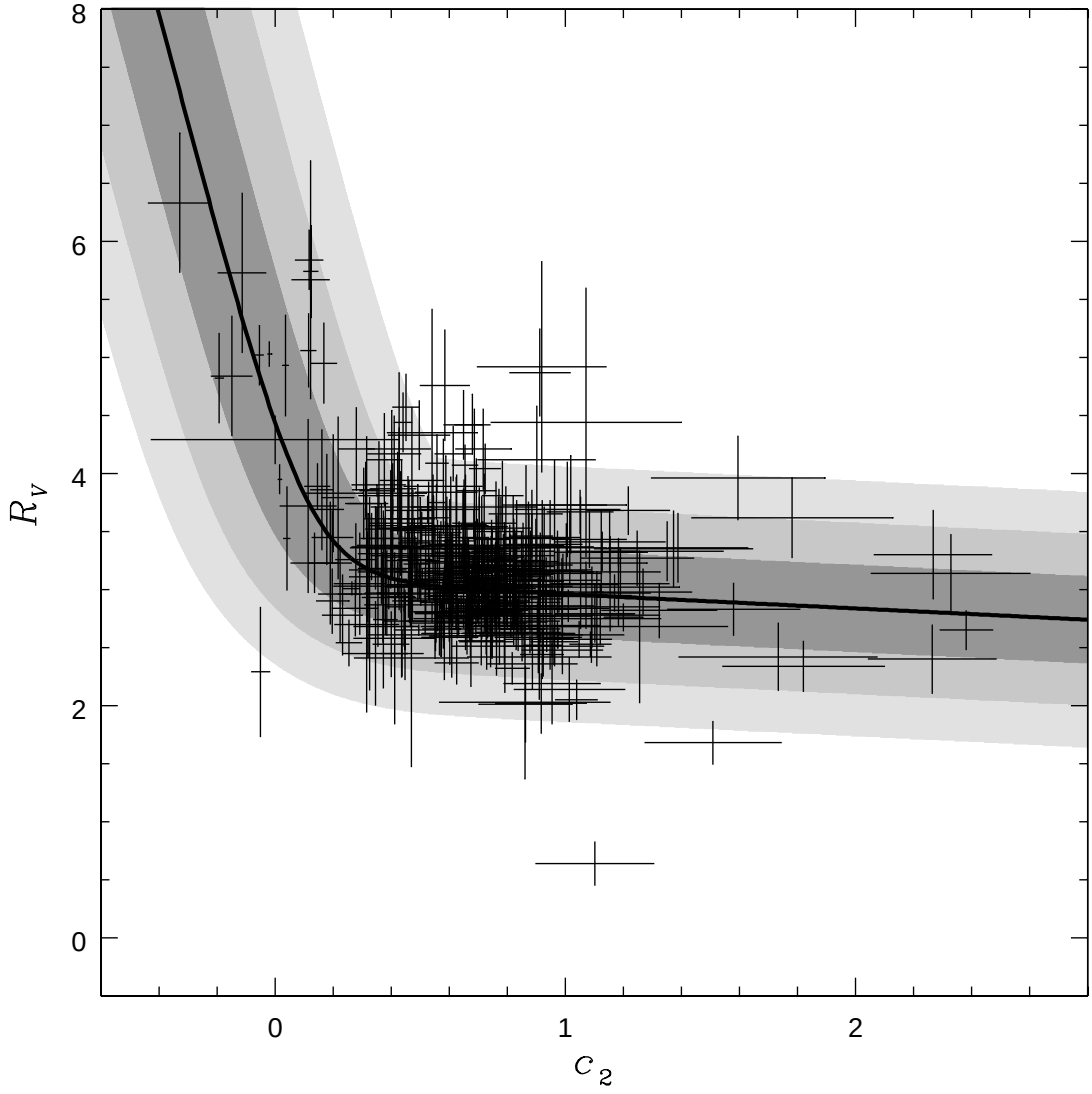

Figure 3.10 Observed R_V vs. c_2 data from Gordon et al. (2003) and Valencic et al. (2004), and fitted smoothly broken linear model distribution. Shaded regions indicate the 1-, 2- and 3σ slop envelopes of the model distribution, using the peak prior values of Table 3.3, and computed using the technique described in Appendix C.

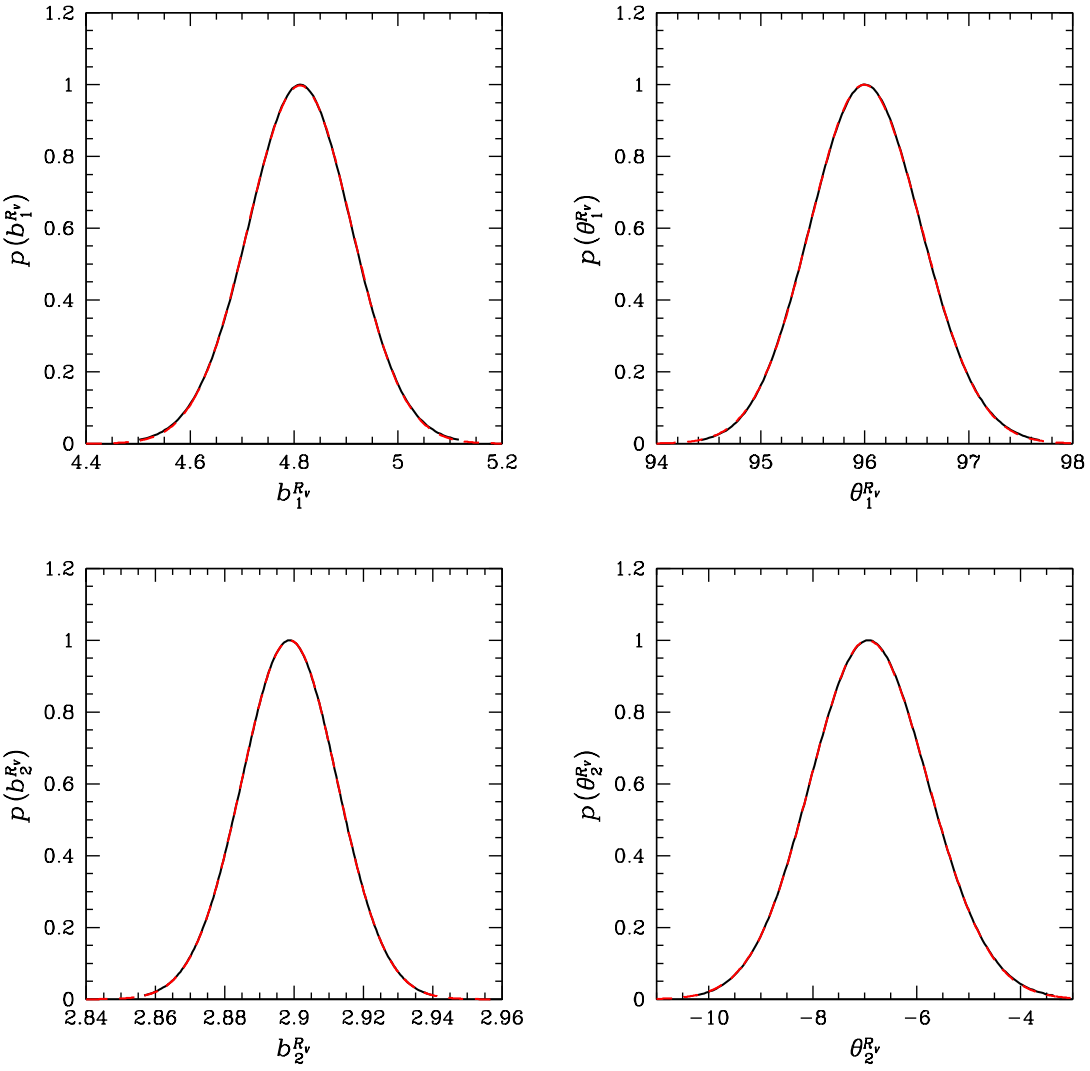

Figure 3.11 Fitted prior probability distributions of R_V vs. c_2 model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.3

Table 3.2. Best-Fit Asymmetric Gaussian Priors and Parameter Values for c_1 vs. c_2 Correlative Model Distribution

Parameter	Peak/ Value	$+1\sigma$ Width	-1σ Width
b^{c_1}	-1.5038	$+0.0245$	-0.0246
θ^{c_1}	$106\overset{\circ}{.}953$	$+0.145$	-0.141
$c_2^{p_{c_1}}$	1.2403	$\cdots$	$\cdots$
$\sigma_{c_2}^{c_1}$	0.08720[a]	$\cdots$	$\cdots$
σ_{c_1}	0.29313[a]	$\cdots$	$\cdots$
s_0	0.25767	$\cdots$	$\cdots$
$R^2_{\rm TRF}$	0.985	$\cdots$	$\cdots$

[a]For 2D TRF fits, slop is scale-dependent, and since fits can only be done at a fixed scale, slop can be interpreted only empirically, not physically.

We employed the full 2D TRF statistic to fit smoothly broken linear model distributions to the weighted R_V vs. c_2 data set, and to find the optimal scale for the fit (§2.5). At any fixed scale, the best-fit extrinsic scatter, or slop, in the x- and y dimensions is described by the parameters $(\sigma_{c_2}^{R_V}, \sigma_{R_V})$, and the model distribution is defined as the convolution of a 2D Gaussian with these widths with a smoothly broken linear model curve:

$$R_{V,c}(c_2;\vartheta_m) = \ln\left[e^{b_1^{R_V}+\tan\theta_1^{R_V}\left(c_2-c_2^{p_{1,R_V}}\right)} + e^{b_2^{R_V}+\tan\theta_2^{R_V}\left(c_2-c_2^{p_{2,R_V}}\right)}\right]. \tag{3.34}$$

Through the iterative technique described in §2.5, we determined the optimum scale for the fit, $s_0 = 0.37568$, at which the correlation coefficient $R^2_{\rm TRF} = 0.820$. From a fine-scale march in θ_1 and θ_2 about their best fit values, we determined values of the pivot points $c_2^{p_{1,R_V}} = -0.0708$ and $c_2^{p_{2,R_V}} = 1.4953$. (§2.6.6). We performed error bar marches (§2.6.5) to determine the projected probability distributions of parameters $b_1^{R_V}$, $\theta_1^{R_V}$,

Table 3.3. Best-Fit Asymmetric Gaussian Priors and Parameter Values for R_V vs. c_2 Correlative Model Distribution

Parameter	Peak/ Value	$+1\sigma$ Width	-1σ Width
$b_1^{R_V}$	4.8118	+0.0990	−0.1004
$\theta_1^{R_V}$	95.°995	+0.548	−0.522
$c_2^{p_{1,R_V}}$	−0.0708	···	···
$b_2^{R_V}$	2.8987	+0.0138	−0.0139
$\theta_2^{R_V}$	−6.°945	+1.165	−1.104
$c_2^{p_{2,R_V}}$	1.4953	···	···
$\sigma_{c_2}^{R_V}$	0.15495[a]	···	···
σ_{R_V}	0.36246[a]	···	···
s_0	0.37568	···	···
R^2_{TRF}	0.820	···	···

[a]For 2D TRF fits, slop is scale-dependent, and since fits can only be done at a fixed scale, slop can be interpreted only empirically, not physically.

$b_2^{R_V}$ and $\theta_2^{R_V}$. Figure 3.9 shows our fitted $p(b_1^{R_V})$, $p(\theta_1^{R_V})$, $p(b_2^{R_V})$ and $p(\theta_2^{R_V})$. To obtain approximate analytic priors on these parameters, we fit asymmetric Gaussian functions to those distributions; these asymmetric Gaussian prior peaks and widths are given in Table 3.3. The model distribution plotted in Figure 3.10 was computed using the peak prior model parameter values of Table 3.3; the 1-, 2- and 3σ envelopes were computed using the technique described in Appendix C.

As with the c_1 vs. c_2 model, the slops $\sigma_{c_2}^{R_V}$ and σ_{R_V} are held fixed at their best-fit values at the optimum scale s_0, and the value of R_V employed in our GRB fits is given by:

$$R_V = R_{V_c}(c_2 - \delta_{c_2}^{R_V}; b_1^{R_V}, \theta_1^{R_V}, b_2^{R_V}, \theta_2^{R_V}) + \delta_{R_V} \,, \tag{3.35}$$

where the curve $R_{V,c}(c_2; b_1^{R_V}, \theta_1^{R_V}, b_2^{R_V}, \theta_2^{R_V})$ is given by Equation 3.34, and where $\delta_{c_2}^{R_V}$ and δ_{R_V} are dummy model parameters constrained by zero-mean Gaussian priors with widths $\sigma_{c_2}^{R_V}$ and σ_{R_V}, i.e., $p(\delta_{c_2}^{R_V}) = G(c_2; 0, \sigma_{c_2}^{R_V})$ and $p(\delta_{R_V}) = G(R_V; 0, \sigma_{R_V})$.

3.3.4.3 Priors on BH vs. c_2

The UV bump height BH $\equiv c_3/\gamma^2$ is also observed to be loosely correlated with c_2: Milky Way lines of sight with moderate values of c_2 tend to exhibit a UV bump, while it tends to be absent in lines of sight with very low or very high values of c_2, which are typical of young star-forming regions like the Orion Nebula and highly processed dust in the Magellanic Clouds, respectively. The data are plotted in Figure 3.12.

We employed the full 2D TRF statistic to fit smoothly broken linear model distributions to the weighted BH vs. c_2 data set, and to find the optimal scale for the fit (§2.5). At any fixed scale, the best-fit extrinsic scatter, or slop, in the x- and y dimensions is described by the parameters $(\sigma_{c_2}^{\rm BH}, \sigma_{\rm BH})$, and the model distribution is defined as the convolution of a 2D Gaussian with these widths with a smoothly broken linear model

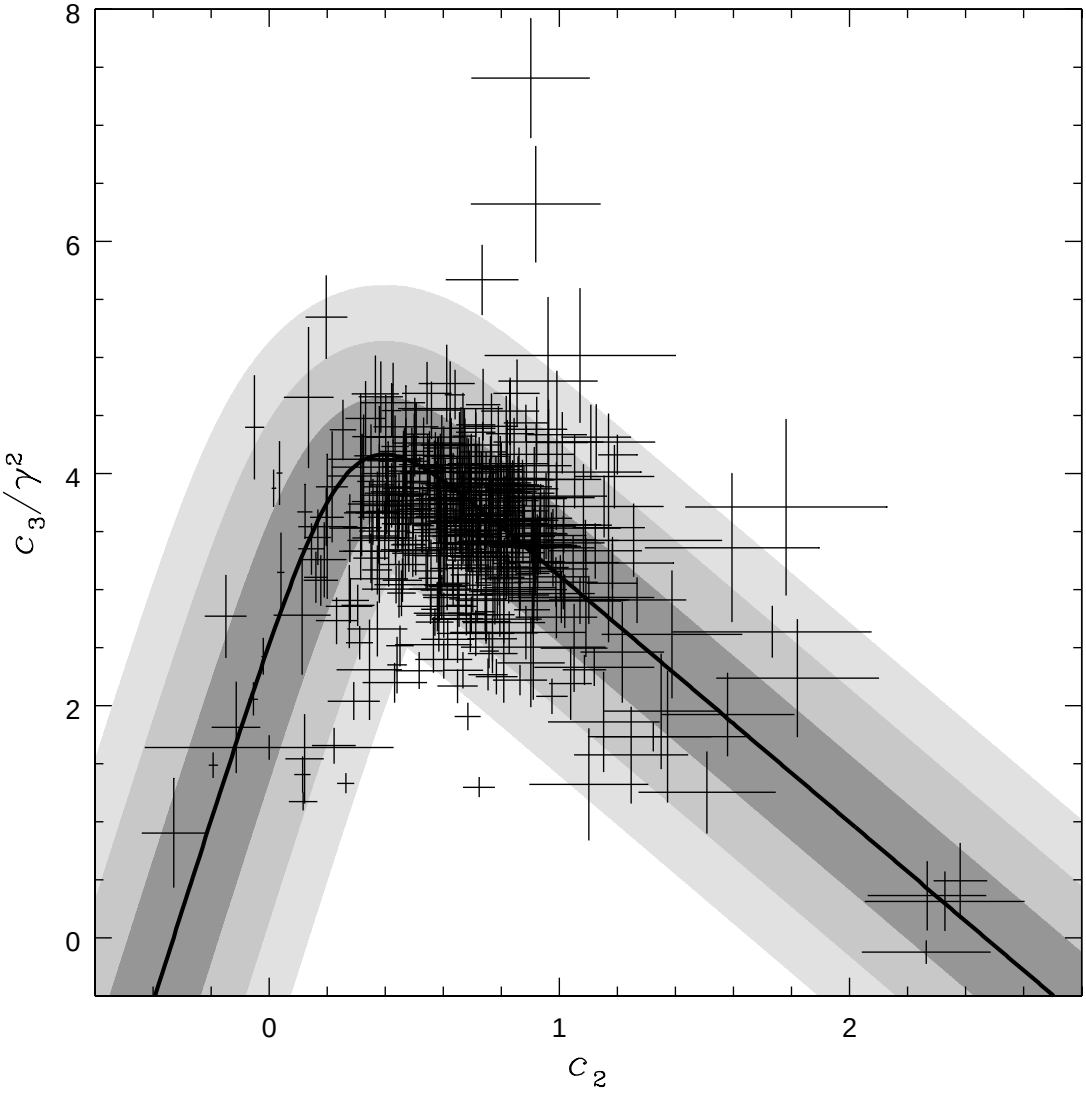

Figure 3.12 Observed BH vs. c_2 data from Gordon et al. (2003) and Valencic et al. (2004), and fitted smoothly broken linear model distribution. Shaded regions indicate the 1-, 2- and 3σ slop envelopes of the model distribution, using the peak prior values of Table 3.4, and computed using the technique described in Appendix C.

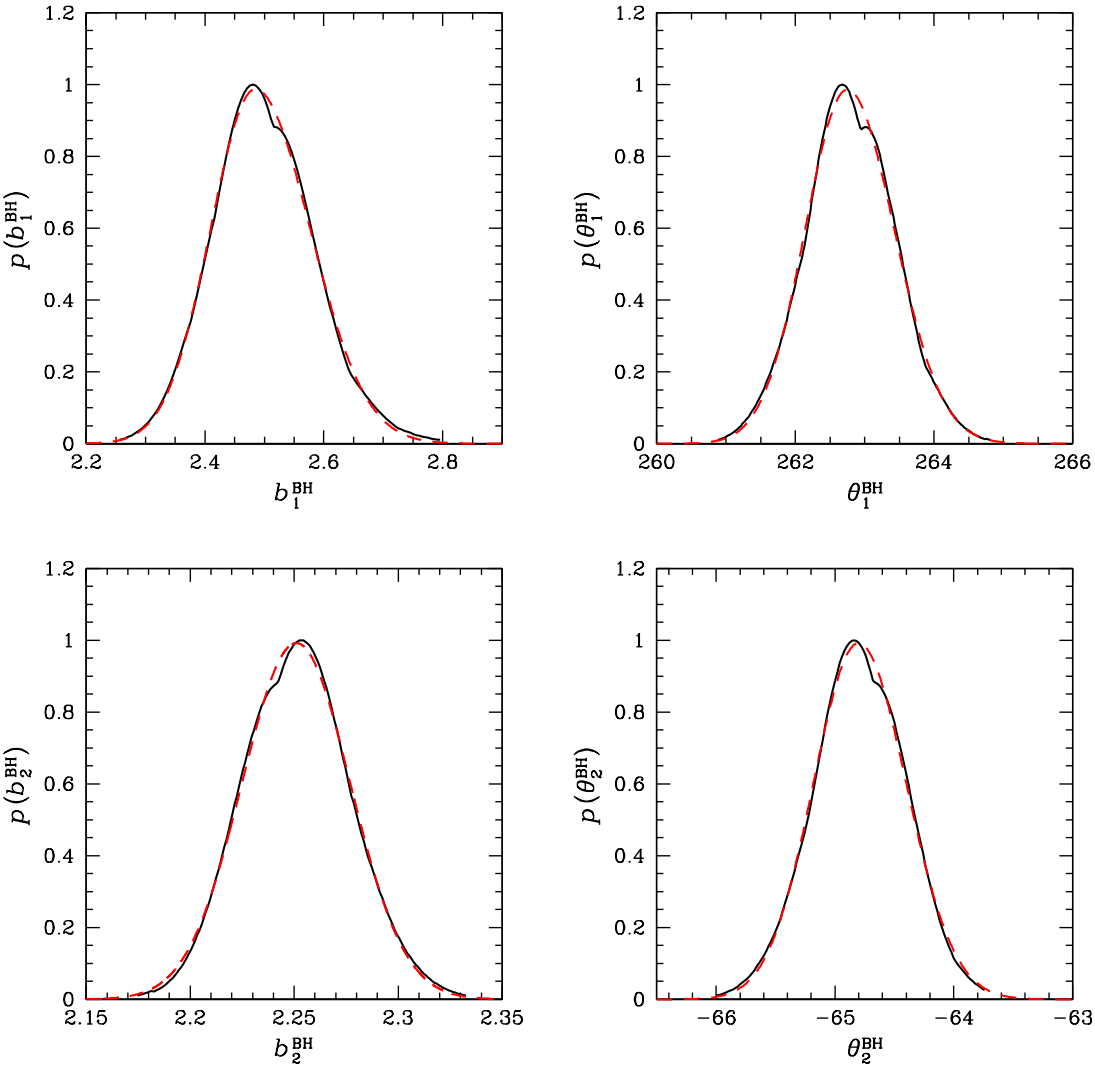

Figure 3.13 Fitted prior probability distributions of BH vs. c_2 model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.4

curve:

$$\mathrm{BH}_c(c_2;\vartheta_m) = -\ln\left[e^{-b_1^{\mathrm{BH}}-\tan\theta_1^{\mathrm{BH}}\left(c_2-c_2^{p_1,\mathrm{BH}}\right)} + e^{-b_2^{\mathrm{BH}}-\tan\theta_2^{\mathrm{BH}}\left(c_2-c_2^{p_2,\mathrm{BH}}\right)}\right] . \qquad (3.36)$$

Through the iterative technique described in §2.5, we determined the optimum scale for the fit, $s_0 = 0.32825$, at which the correlation coefficient $R^2_{\mathrm{TRF}} = 0.769$. From a fine-scale march in θ_1^{BH} and θ_2^{BH} about their best fit values, we determined values of the pivot points $c_2^{p_1,\mathrm{BH}} = -0.0143$ and $c_2^{p_2,\mathrm{BH}} = 1.4087$. (§2.6.6). We performed error bar marches (§2.6.5) to determine the projected probability distributions of parameters b_1^{BH}, θ_1^{BH}, b_2^{BH} and θ_2^{BH}. Figure 3.13 shows our fitted $p(b_1^{\mathrm{BH}})$, $p(\theta_1^{\mathrm{BH}})$, $p(b_2^{\mathrm{BH}})$ and $p(\theta_2^{\mathrm{BH}})$. To obtain approximate analytic priors on these parameters, we fit asymmetric Gaussian functions to those distributions; these asymmetric Gaussian prior peaks and widths are given in Table 3.4. The model distribution plotted in Figure 3.12 was computed using the peak prior model parameter values of Table 3.4; the 1-, 2- and 3σ envelopes were computed using the technique described in Appendix C.

As with the c_1 and R_V vs. c_2 models, the slops are held fixed at their best-fit values at the optimum scale s_0, and the value of BH employed in our GRB fits is given by:

$$\mathrm{BH} = \mathrm{BH}_c(c_2 - \delta_{c_2}^{\mathrm{BH}}; b_1^{\mathrm{BH}}, \theta_1^{\mathrm{BH}}, b_2^{\mathrm{BH}}, \theta_2^{\mathrm{BH}}) + \delta_{\mathrm{BH}} , \qquad (3.37)$$

where the curve $\mathrm{BH}_c(c_2; b_1^{\mathrm{BH}}, \theta_1^{\mathrm{BH}}, b_2^{\mathrm{BH}}, \theta_2^{\mathrm{BH}})$ is given by Equation 3.36, and where $\delta_{c_2}^{\mathrm{BH}}$ and δ_{BH} are dummy model parameters constrained by zero-mean Gaussian priors with widths $\sigma_{c_2}^{\mathrm{BH}}$ and σ_{BH}, i.e., $p(\delta_{c_2}^{\mathrm{BH}}) = G(c_2; 0, \sigma_{c_2}^{\mathrm{BH}})$ and $p(\delta_{\mathrm{BH}}) = G(\mathrm{BH}; 0, \sigma_{\mathrm{BH}})$.

3.3.4.4 Priors on x_0 and γ

The frequency of the peak of the UV bump, described by the parameter x_0, and the bump width, described by the parameter γ, are not observed to be significantly corre-

Table 3.4. Best-Fit Asymmetric Gaussian Priors and Parameter Values for BH vs. c_2 Correlative Model Distribution

Parameter	Peak/ Value	$+1\sigma$ Width	-1σ Width
b_1^{BH}	2.4845	+0.0925	−0.0750
θ_1^{BH}	262.°749	+0.677	−0.607
$c_2^{p_1,\mathrm{BH}}$	−0.0143	···	···
b_2^{BH}	2.2511	+0.0259	−0.0262
θ_2^{BH}	−64.°803	+0.402	−0.379
$c_2^{p_2,\mathrm{BH}}$	1.4087	···	···
$\sigma_{c_2}^{\mathrm{BH}}$	0.14246[a]	···	···
σ_{BH}	0.48315[a]	···	···
s_0	0.32825	···	···
R^2_{TRF}	0.769	···	···

[a]For 2D TRF fits, slop is scale-dependent, and since fits can only be done at a fixed scale, slop can be interpreted only empirically, not physically.

lated with any of the other FM extinction model parameters. Figures 3.14 & 3.17 show the x_0 and γ data plotted versus the parameter c_2. We excluded a priori four stars from the SMC Bar sample of Gordon et al. (2003), which were reported with fiducial values of $x_0 = 4.6 \pm 0.0$ and $\gamma = 1.00 \pm 0.31$.[5]

We fit model distributions to these data of the form $x_{0,c}(c_2; \vartheta_m) = \text{constant} = b_{x_0}$ and $\gamma_c(c_2; \vartheta_m) = \text{constant} = b_\gamma$, convolved with asymmetric Gaussian slop with widths $\sigma_{x_0\pm}$ and $\sigma_{\gamma\pm}$, respectively. Since, in Figures 3.14 & 3.17, c_2 is being used only as an index for plotting, analogous to what we did in Figure 3.2 for $\log R_V^{\rm MW}$, in our fits to these data the 2D TRF statistic reduces to a 1D χ^2-like statistic (§2.4.2), and the joint probability and likelihood functions take the same 1D forms as do those for the $\log R_V^{\rm MW}$ model distribution (§3.3.2). In Figures 3.14 & 3.17, we plot x_0 and γ versus c_2 instead of a randomized index only to aid in visualizing the weighting of the data points (Figure 3.7).

Since these fits employ a one-dimensional statistic, the fitted slops are scale-invariant, and so, just as for $R_V^{\rm MW}$, we must compute and apply the projected probability distributions of both the constant and the slop parameters in our model implementation. We performed error bar marches (§2.6.5) to determine the projected probability distributions of parameters b_{x_0}, σ_{x_0+}, σ_{x_0-}, b_γ, $\sigma_{\gamma+}$ and $\sigma_{\gamma-}$. Figure 3.15 shows our fitted $p(b_{x_0})$, $p(\sigma_{x_0+})$, and $p(\sigma_{x_0-})$, and Figure 3.18 shows our fitted $p(b_\gamma)$, $p(\sigma_{\gamma+})$, and $p(\sigma_{\gamma-})$. To obtain approximate analytic priors on these parameters, we fit asymmetric Gaussian functions to those distributions; these asymmetric Gaussian prior peaks and widths are given in Table 3.5. The model distributions plotted in Figure 3.14 & 3.17 were computed using the peak prior model parameter values of Table 3.5; the 1-, 2- and 3σ envelopes were computed using the technique described in Appendix C.

[5]This is presumably because these parameters were held fixed in their fits to these spectra; we do include these stars in our UV bump height analysis, however, since BH $= c_3/\gamma^2$, and a statistically valid value of, and uncertainty in, c_3 is reported; furthermore, these SMC Bar lines of sight provide important "anchor points" at high c_2 in that analysis.

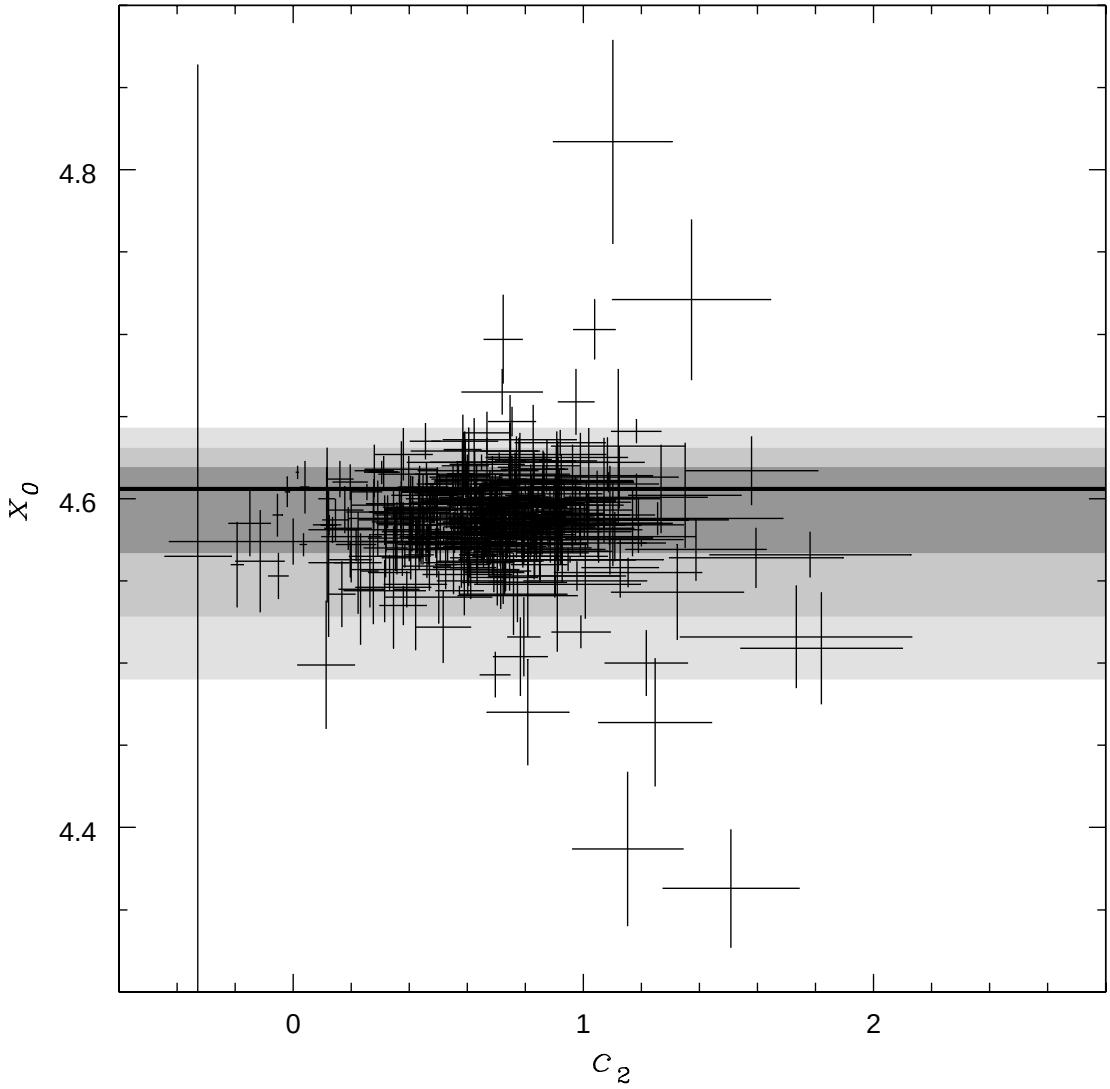

Figure 3.14 Observed x_0 vs. c_2 data from Gordon et al. (2003) and Valencic et al. (2004), and fitted asymmetric Gaussian model distribution. Shaded regions indicate the 1-, 2- and 3σ slop envelopes of the model distribution, using the peak prior values of Table 3.5.

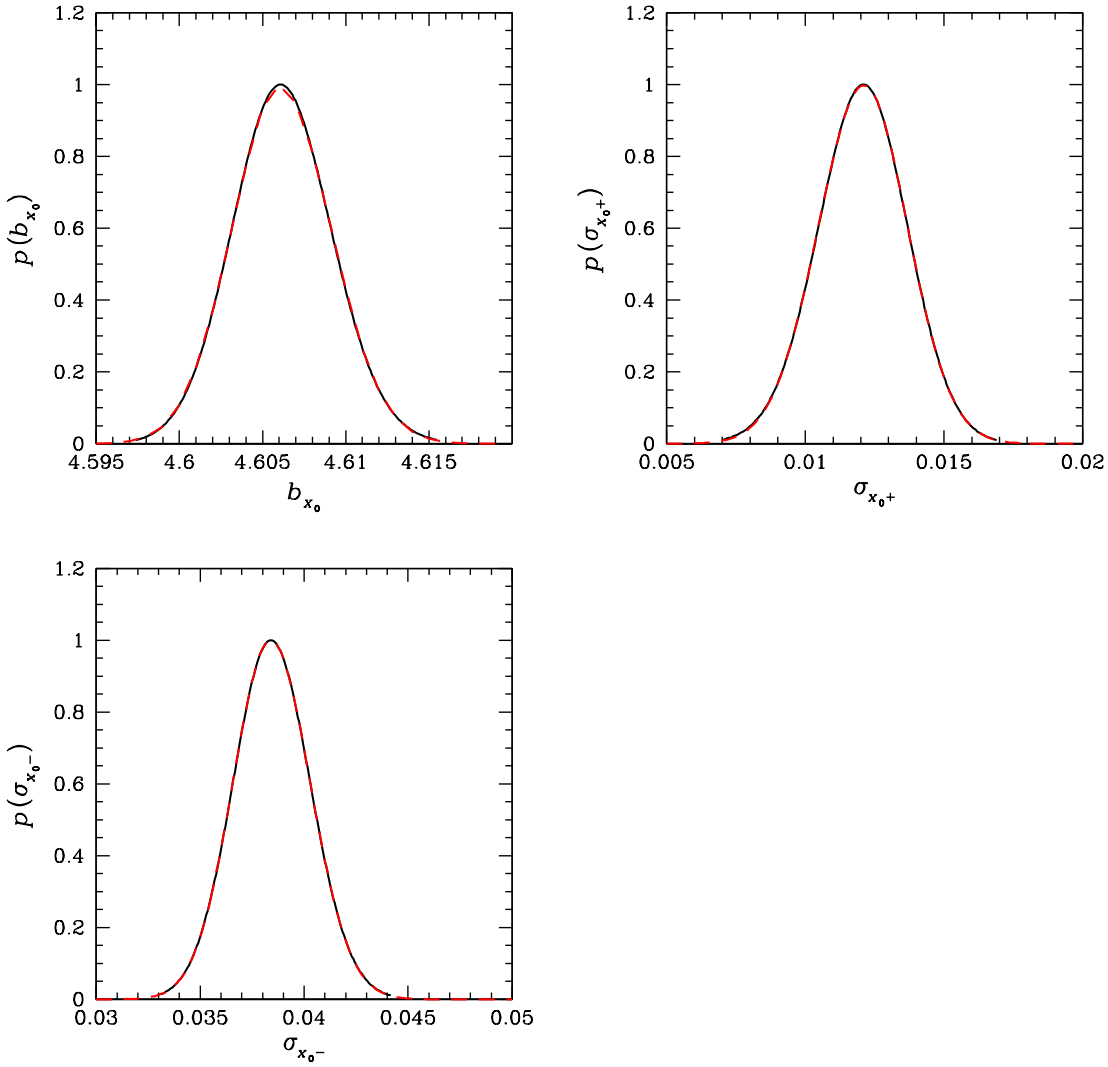

Figure 3.15 Fitted prior probability distributions of x_0 model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.5

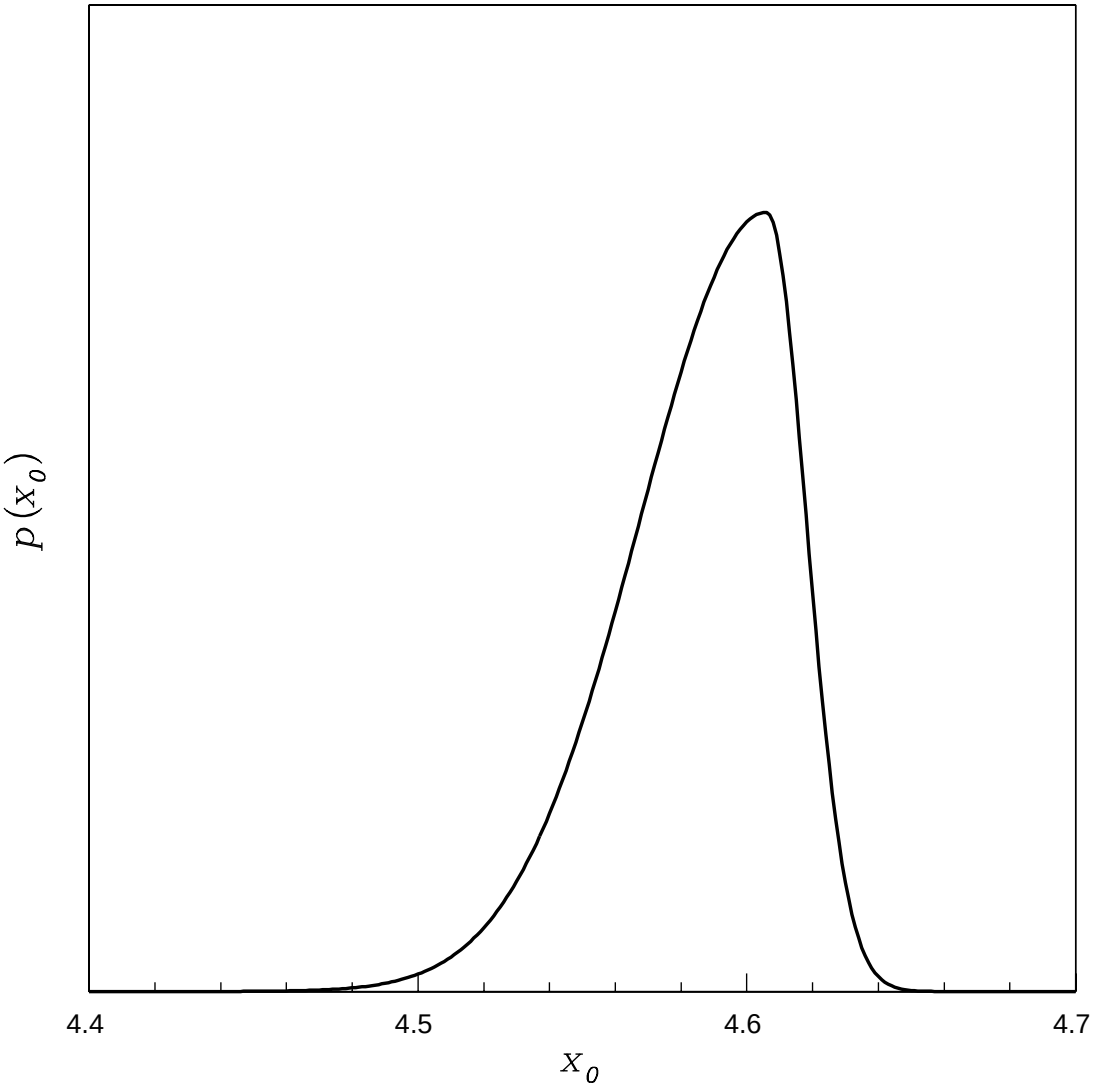

Figure 3.16 Asymmetric Gaussian model distribution fit to the observed distribution of the extinction parameter x_0, using the peak prior parameter values of Table 3.5.

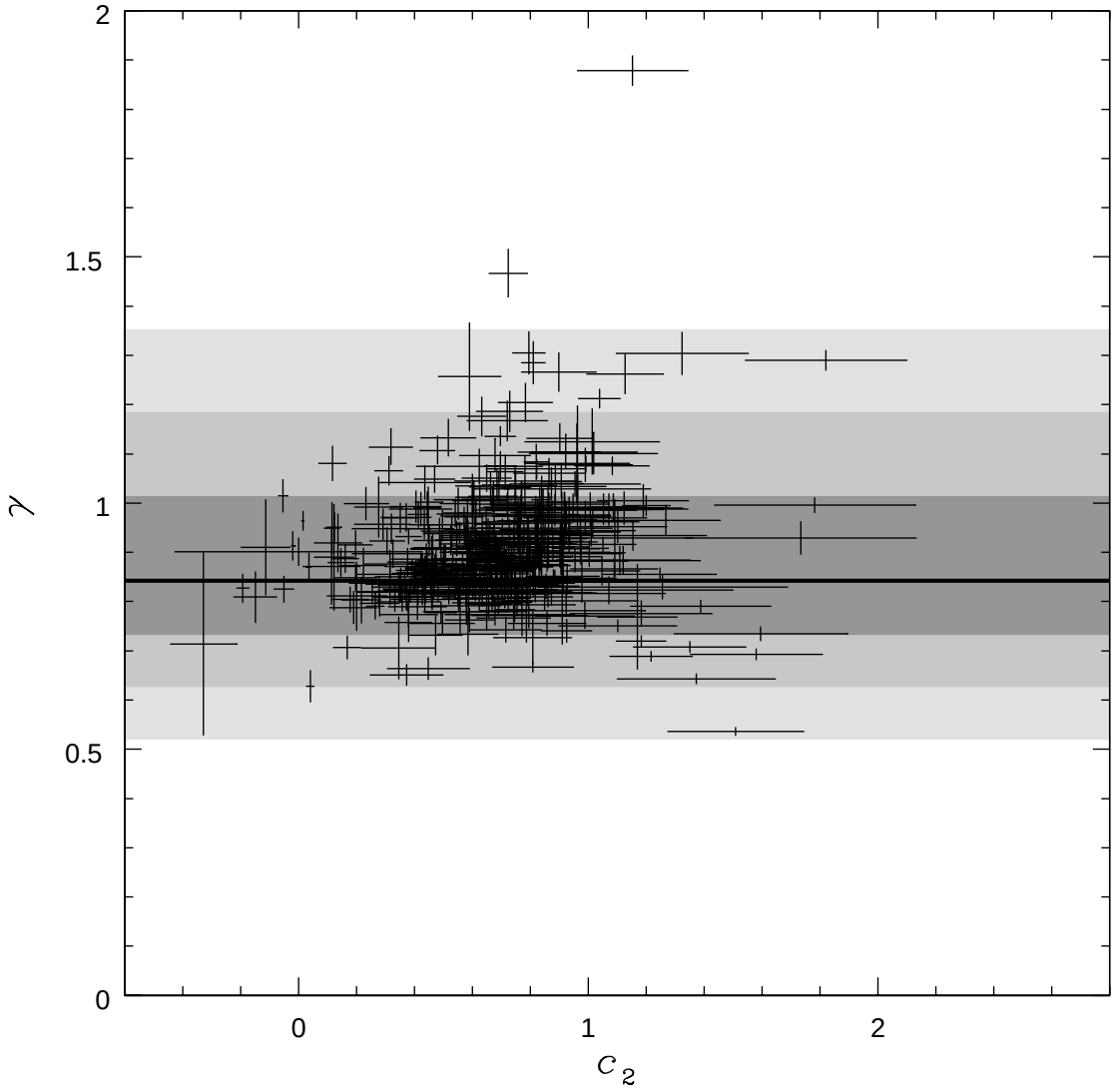

Figure 3.17 Observed γ vs. c_2 data from Gordon et al. (2003) and Valencic et al. (2004), and fitted asymmetric Gaussian model distribution. Shaded regions indicate the 1-, 2- and 3σ slop envelopes of the model distribution, using the peak prior values of Table 3.5.

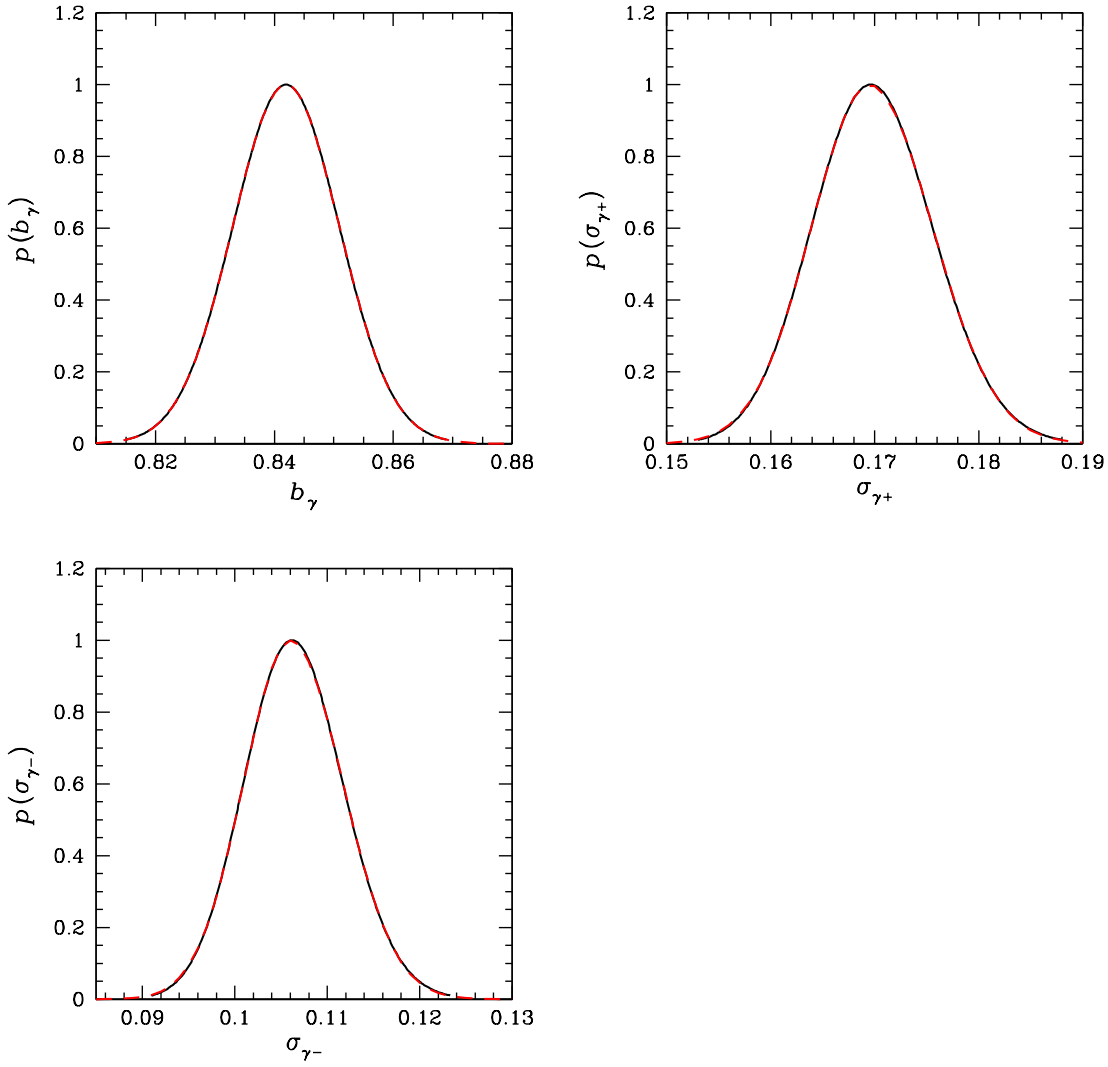

Figure 3.18 Fitted prior probability distributions of γ model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.5

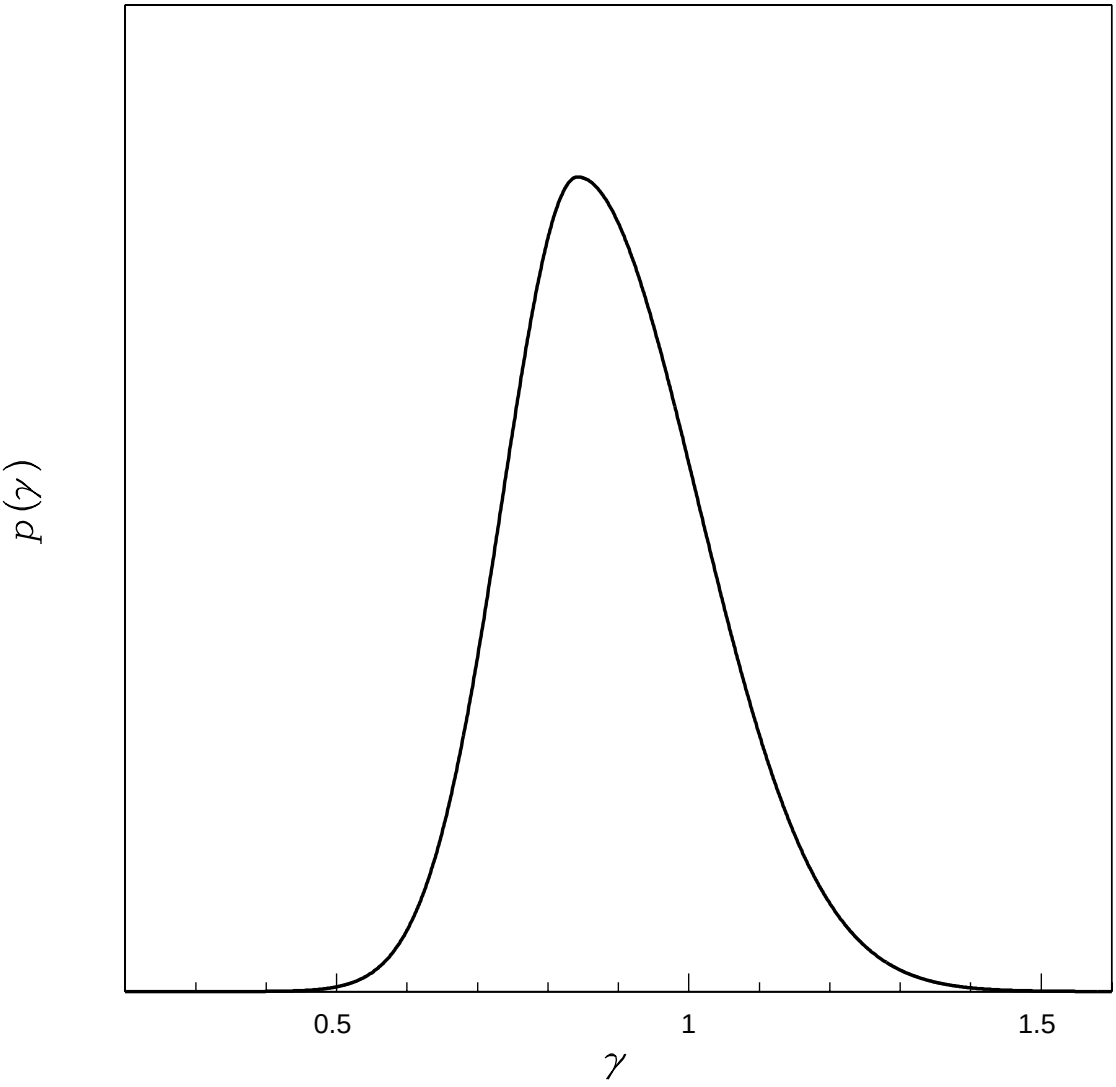

Figure 3.19 Asymmetric Gaussian model distribution fit to the observed distribution of the extinction parameter γ, using the peak prior parameter values of Table 3.5.

Table 3.5. Best-Fit Asymmetric Gaussian Priors for x_0 and γ Probability Distributions

Parameter	Peak	$+1\sigma$ Width	-1σ Width
b_{x_0}	4.60604	+0.00305	−0.00286
σ_{x_0+}	0.01212	+0.00157	−0.00165
σ_{x_0-}	0.03839	+0.00189	−0.00181
b_γ	0.84195	+0.00897	−0.00895
$\sigma_{\gamma+}$	0.16949	+0.00605	−0.00555
$\sigma_{\gamma-}$	0.10605	+0.00562	−0.00508

In our GRB modeling, we fit to the parameters b_{x_0}, σ_{x_0+}, σ_{x_0-}, b_γ, $\sigma_{\gamma+}$ and $\sigma_{\gamma-}$, each constrained by its respective asymmetric Gaussian prior, given in Table 3.5. The actual values of x_0 and γ that we employ in our GRB modeling are:

$$\begin{aligned} x_0 &= b_{x_0} + \delta_{x_0}\,, \\ \gamma &= b_\gamma + \delta_\gamma\,, \end{aligned} \tag{3.38}$$

where δ_{x_0} and δ_γ are dummy model parameters, each constrained by zero-mean asymmetric Gaussian priors with $\pm1\sigma$ widths given by the prior-constrained slop parameters $\sigma_{x_0\pm}$ and $\sigma_{\gamma\pm}$, respectively, i.e., $p(\delta_{x_0}) = G_A(x_0; 0, \sigma_{x_0\pm})$ and $p(\delta_\gamma) = G_A(\gamma; 0, \sigma_{\gamma\pm})$.

3.4 Absorption Models

3.4.1 Damped Lyα Absorber

Our model includes a damped Lyα absorption line redshifted from the GRB source frame, centered at $\nu^{\rm obs}_{{\rm Ly}\alpha} = 2.46605 \times 10^{15}/(1 + z_{\rm GRB})$ Hz. The line profile is a function of the neutral hydrogen column density N_H and the source redshift $z_{\rm GRB}$ only (see, e.g.,

Totani et al. (2006)):

$$\Delta \log F_\nu^{\mathrm{Ly}\alpha} = \frac{-2.9979 \times 10^6 \left(\frac{\nu(1+z_{\mathrm{GRB}})}{\nu_{\mathrm{Ly}\alpha}}\right)^4 N_H}{39.4784 \left(\nu(1+z_{\mathrm{GRB}}) - \nu_{\mathrm{Ly}\alpha}\right)^2 + 9.78275 \times 10^{16} \left(\frac{\nu(1+z_{\mathrm{GRB}})}{\nu_{\mathrm{Ly}\alpha}}\right)^6}, \quad (3.39)$$

where ν is in Hz and N_H in cm^{-2}. We can fit to N_H as a free model parameter, or, if reliable optical spectroscopic data are available, constrain it by a prior.[6] Modeling the Lyα absorption line is also important because it can coincide with the IGM Gunn-Peterson trough for high-z sources (see §3.4.2); in this case, both must be modeled to correctly interpret either.

We do not include Lyβ, Lyγ, etc. absorption lines in our model. The absorption lines of the higher-level Lyman series are expected to have lower equivalent widths than Lyα, making them less likely to impact photometric measurements, and they cannot coincide with the Gunn-Peterson trough. But, in principle, they could be included, e.g., if one chose to fit models to spectroscopic data. However, we always assume total absorption at frequencies blueward of the Lyman limit in the source frame, $\nu > 3.288037 \times 10^{15}$ Hz$/(1 + z_{\mathrm{GRB}})$.

3.4.2 IGM Absorption: The Lyα Forest and Gunn-Peterson Trough

To model Lyα absorption in the intergalactic medium, we fit an empirical model to transmission T versus absorber redshift z_{abs} based on observed flux deficits for 64 QSOs, measured in binned regions of width Δz_{abs} of their spectra blueward of Lyα and redward of Lyβ in the QSO source frame. Our data set combines 50 mid-to-high

[6]We note that it is more common to have a prior measurement of N_H^{X} based on fits to X-ray spectra. However, these must be used with caution in modeling Lyα absorption; the majority of X-ray absorption is actually due to metals, and N_H^{X} is inferred assuming solar metal abundance ratios (Morrison & McCammon, 1983) that may not apply to all bursts, especially to those at high redshifts.

redshift QSOs analyzed by Songaila (2004, hereafter S04), and 19 high-redshift QSOs analyzed by Fan et al. (2006, hereafter F06). Five QSOs in S04 overlapped with F06 – for these, we used the F06 data, which were tabulated in a more user-friendly format (The S04 data was extracted by hand from Figure 1a of that paper; the data points corresponding to a given quasar were identified by finding regular patterns in plotted absorber redshift). Both papers quote Lyα fractional transmission, T, versus absorber redshift, $z_{\rm abs}$. Observed frequency $\nu_{\rm obs}$ is related to $z_{\rm abs}$ by:

$$\nu_{\rm obs} = \frac{\nu_{\rm Ly\alpha}}{1 + z_{\rm abs}}\,, \tag{3.40}$$

where $\nu_{\rm Ly\alpha} = 2.466028 \times 10^{15}$ Hz is the rest frame frequency of the Lyα transition of atomic hydrogen.

The transmission data of S04 and F06 are quoted in linear T space, with symmetric error bars.[7] We experimented with various empirical parameterizations for the model distribution describing $T(z_{\rm abs})$, in a variety of bases. We ultimately chose to fit a model distribution that is described by a smoothly broken linear function in $\ln[-\ln T(z_{\rm abs})]$ space, where the slop in the T direction appears to be roughly symmetric. Figure 3.20 shows the data transformed into this space. Figure 3.21 shows the data transformed into $\ln T$ space, and Figure 3.22 shows the original data in linear T space.

We have a priori reason to assume that any slop in the data is in the T direction, due to there being varying numbers of discrete Lyα absorbers in different redshift bins along different lines of sight. Moreover, the intrinsic uncertainty in the center of a given redshift bin $\sigma_{z_{\rm abs},n}$, projected into the T dimension, is negligible compared to both the

[7]Actually, the S04 data do not have quoted error bars. In the caption to Figure 1a of that paper is the statement, "When the transmitted flux is less than 0.0015, we have shown it at this nominal value." We therefore assign all S04 data points symmetric intrinsic error bars $\sigma_{T,n} = 0.0015$, and for the 3 data points – presumably non-detections – plotted at $T = 0.0015$, we assign them nominal values $T_n = 0$. We recognize that this is less than ideal, possibly underestimating the intrinsic uncertainties in these data. However, since this is in fact a heavily slop-dominated data set, having precise values for the intrinsic uncertainties is not critical.

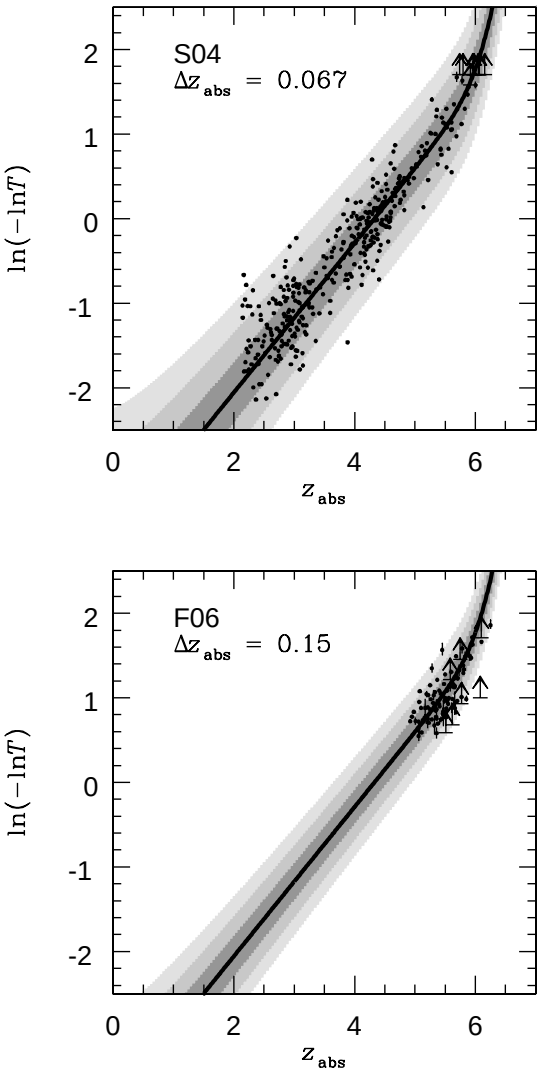

Figure 3.20 Lyα absorption vs. absorber redshift: $\ln(-\ln T)$ vs. $z_{\rm zabs}$. Top: S04 data and model distribution for median binwidth $\Delta z_{\rm abs} = 0.067$. Bottom: F06 data and model distribution for binwidth $\Delta z_{\rm abs} = 0.15$. The 1-, 2- and 3$\sigma$ model distribution envelopes were computed using the peak prior values of Table 3.6.

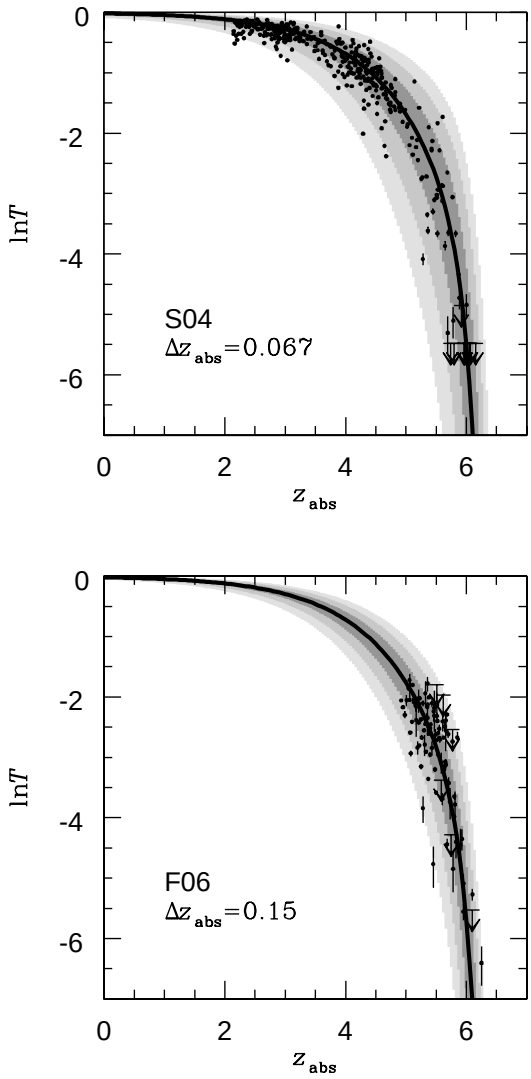

Figure 3.21 Lyα absorption vs. absorber redshift: $\ln T$ vs. $z_{\rm zabs}$. Top: S04 data and transformed model distribution for median binwidth $\Delta z_{\rm abs} = 0.067$. Bottom: F06 data and transformed model distribution for binwidth $\Delta z_{\rm abs} = 0.15$. The 1-, 2- and 3$\sigma$ model distribution envelopes were computed using the peak prior values of Table 3.6, and transformed from $\ln(-\ln T)$ space.

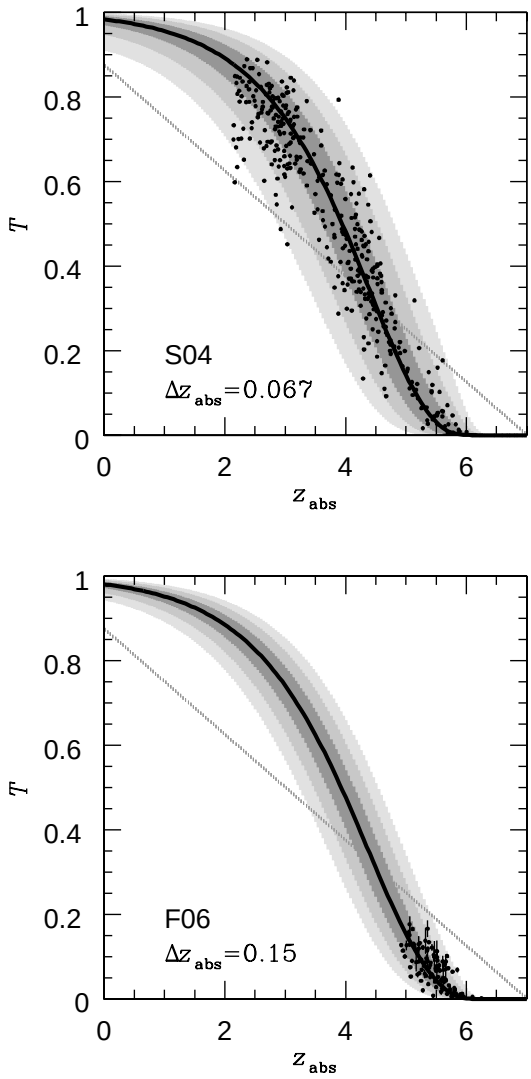

Figure 3.22 Lyα absorption vs. absorber redshift: T vs. $z_{\rm zabs}$. Top: S04 data and transformed model distribution for median binwidth $\Delta z_{\rm abs} = 0.067$. Bottom: F06 data and transformed model distribution for binwidth $\Delta z_{\rm abs} = 0.15$. The 1-, 2- and 3$\sigma$ model distribution envelopes were computed using the peak prior values of Table 3.6, and transformed from $\ln(-\ln T)$ space.

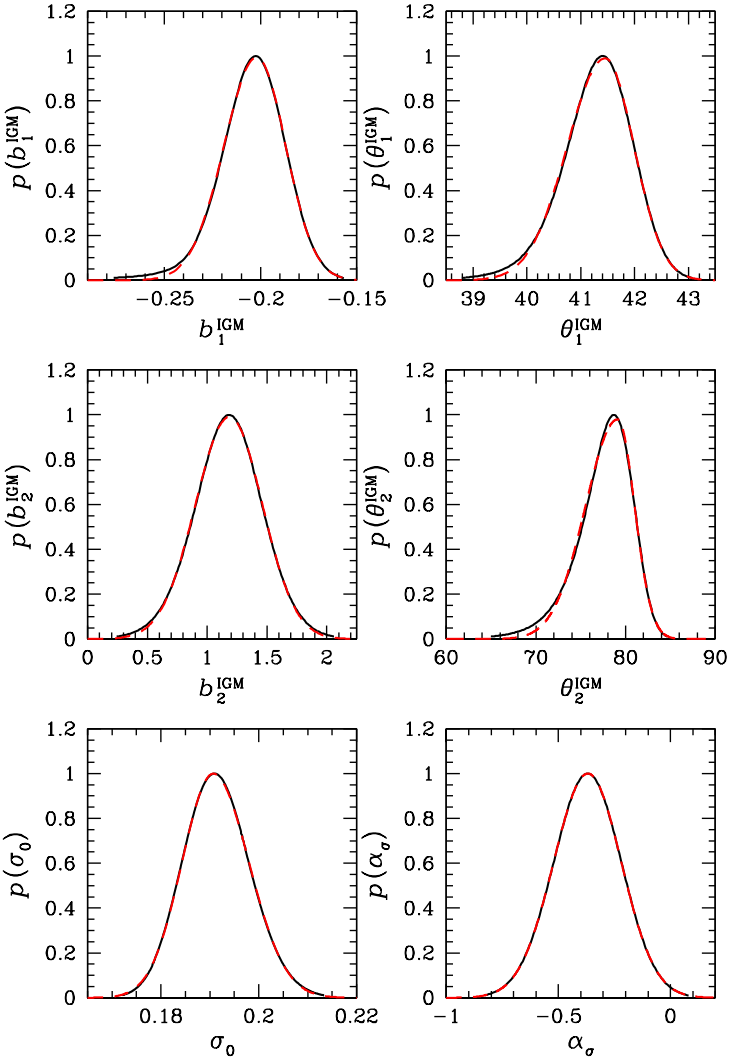

Figure 3.23 Fitted prior probability distributions of IGM $\ln(-\ln T)$ vs. $z_{\rm abs}$ model parameters (solid black curves). Dashed red curves are best-fit asymmetric Gaussian approximations to the actual pdfs. Peaks and $\pm 1\sigma$ widths of these approximations are given in Table 3.6

intrinsic uncertainty in the transmission $\sigma_{T,n}$ and the slop in the data σ_T. We therefore chose to fit our model distribution to the T_n vs. $z_{\mathrm{abs},n}$ data set using a 1D statistic.

In the spirit of "bringing the model to the data", we performed the fit in the linear T space, transforming the symmetric Gaussian model distribution from $\ln(-\ln T)$ space using the technique described in §2.7.1. This particular transformation does not turn out to be amenable to analytic, asymmetric Gaussian approximations. We therefore computed the 1D joint probability integral of the transformed model distribution with the Gaussian intrinsic distribution of each data point directly, using a simple trapezoidal numerical integration algorithm.

In $\ln(-\ln T)$ space, the model distribution is described by the convolution of a symmetric, 1D Gaussian function of width $\sigma_{\ln(-\ln T)}$ with the smoothly broken linear model curve:

$$\ln\left[-\ln T_c(z_{\mathrm{abs}};\vartheta_m)\right] = \ln\left[e^{b_1^{\mathrm{IGM}}+\tan\theta_1^{\mathrm{IGM}}(z_{\mathrm{abs}}-z_1)} + e^{b_2^{\mathrm{IGM}}+\tan\theta_2^{\mathrm{IGM}}(z_{\mathrm{abs}}-z_2)}\right] . \qquad (3.41)$$

However, in this case, we do *not* assume that the slop $\sigma_{\ln(-\ln T)}$ is constant across the entire data set. The slop is expected to be primarily a function of the number of discrete Lyα absorbers there are in a given absorber redshift bin, of width Δz_{abs}. The more discrete absorbers that were averaged in the spectra to obtain a given transmission measurement, the smaller the expected slop. There are two trends we might expect: 1) The number of absorbers at a given redshift increases linearly with redshift binwidth, Δz_{abs}; and 2) The number of absorbers per unit binwidth should increase with increasing redshift since, in the early universe, neutral hydrogen clouds in the IGM were both more numerous and more closely packed together. We account for the first effect by scaling the slop for a given data point by the binwidth of that point, assuming Poisson statistics, i.e.:

$$\sigma_{-\ln(\ln T)} \propto \left(\Delta z_{\mathrm{abs}}\right)^{-1/2} . \qquad (3.42)$$

We account for the second effect empirically, scaling the slop as a power law in redshift:

$$\sigma_{-\ln(\ln T)} \propto (1 + z_{\rm abs})^{\alpha_\sigma} \, . \tag{3.43}$$

We thus assume normally distributed slop in $\ln(-\ln T)$ of the form:

$$\sigma_{\ln(-\ln T)}(z_{\rm abs}) = \sigma_0 \left(\frac{1 + z_{\rm abs}}{1 + z_0} \right)^{\alpha_\sigma} \left(\frac{\Delta z_{\rm abs}}{\Delta z_0} \right)^{-\frac{1}{2}} \, . \tag{3.44}$$

The reference "pivot point" redshifts $z_1 = 4.10$ and $z_2 = 6.15$ in Equation 3.41, and the reference redshift $z_0 = 4.23$ in Equation 3.44 were computed using the method described in §2.6.6 to minimize correlations among the fitted IGM model parameters. The reference binwidth $\Delta z_0 = 0.15$ is an arbitrary choice, chosen to match the binwidth of the F06 data set. We performed error bar marches (§2.6.5) to determine the projected probability distributions of parameters $b_1^{\rm IGM}$, $\theta_1^{\rm IGM}$, $b_2^{\rm IGM}$, $\theta_2^{\rm IGM}$, σ_0 and α_σ. Figure 3.23 shows our fitted $p(b_1^{\rm IGM})$, $p(\theta_1^{\rm IGM})$, $p(b_2^{\rm IGM})$, $p(\theta_2^{\rm IGM})$, $p(\sigma_0)$ and $p(\alpha_\sigma)$. To obtain approximate analytic priors on these parameters, we fit asymmetric Gaussian functions to those distributions; these asymmetric Gaussian prior peaks and widths are given in Table 3.6. The model distributions plotted in Figure 3.20, 3.21 & 3.22 were computed using the peak prior model parameter values of Table 3.6; the 1-, 2- and 3σ envelopes were computed using the technique described in Appendix C. Figure 3.24 shows sample model GRB spectra, with this IGM absorption model included, for a range of source redshifts.

In GRB afterglow fits, we model IGM absorption by fitting to the secondary parameters $b_1^{\rm IGM}$, $\theta_1^{\rm IGM}$, $b_2^{\rm IGM}$, $\theta_2^{\rm IGM}$, σ_0 and α_σ, each of which is constrained the asymmetric Gaussian priors of Table 3.6. For each photometric filter f, we compute the filter response-weighted mean absorber redshift $\overline{z}_f$ and effective binwidth $\overline{\Delta z}_f$, from the portion of the filter that is blueward of Lyα and redward of the Lyman limit in the source

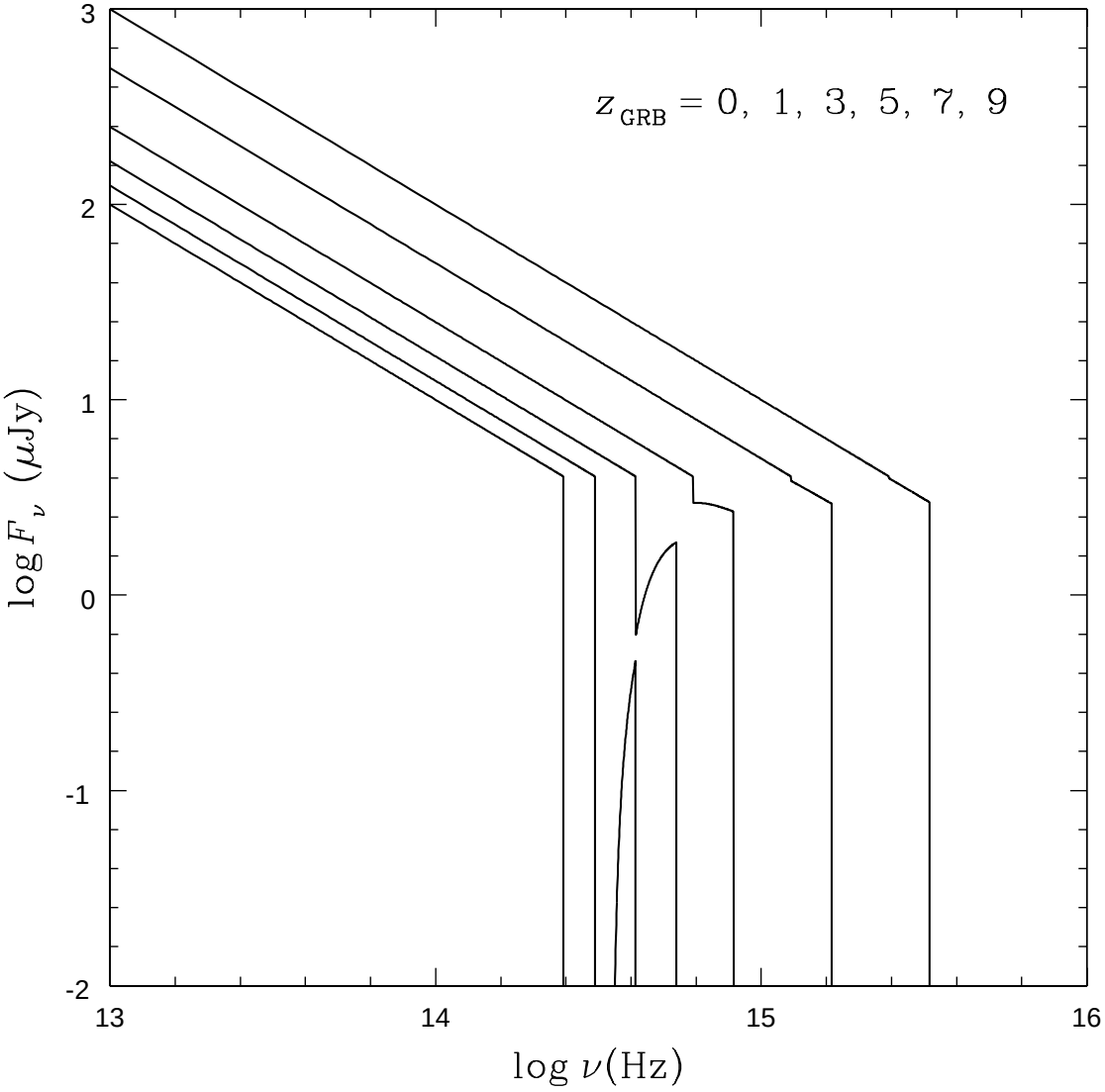

Figure 3.24 Model spectral flux densities for a GRB at at various redshifts, with an intrinsic power-law spectrum $F_\nu \propto \nu^{-1}$. Curves show model spectra, including the Lyα forest/Gunn-Peterson trough, for $z_{\rm GRB} = 0, 1, 3, 5, 7, 9$. Note that total absorption is assumed at frequencies greater than the Lyman limit, $\log \nu_{\rm LL} = 15.52$ in the source frame of the GRB.

Table 3.6. Best-Fit Asymmetric Gaussian Priors and Parameter Values for IGM $\ln(-\ln T)$ vs. $z_{\rm abs}$ Correlative Model Distribution

Parameter	Peak/ Value	$+1\sigma$ Width	-1σ Width
$b_1^{\rm IGM}$	-0.20184	$+0.01490$	-0.01706
$\theta_1^{\rm IGM}$	$41\overset{\circ}{.}4538$	$+0.5342$	-0.7043
z_1	4.10	$\cdots$	$\cdots$
$b_2^{\rm IGM}$	1.18711	$+0.27507$	-0.28488
$\theta_2^{\rm IGM}$	$79\overset{\circ}{.}1200$	$+1.8779$	-3.5870
z_2	6.15	$\cdots$	$\cdots$
σ_0	0.190677	$+0.007318$	-0.006359
α_σ	-0.3682	$+0.1488$	-0.1526
z_0	4.23	$\cdots$	$\cdots$
Δz_0	0.15	$\cdots$	$\cdots$

frame (i.e., $\overline{z}_f$ and $\overline{\Delta z}_f$ are the effective mean absorber redshift and binwidth of that part of the filter that overlaps the Lyα forest.) We then introduce a dummy parameter $\delta_f^{\rm IGM}$ in $\ln(-\ln T)$ for each filter, constrained by a zero-mean Gaussian prior of width:

$$\sigma_f^{\rm IGM} = \left(\frac{1+\overline{z}_f}{1+z_0}\right)^{\alpha_\sigma} \left(\frac{\overline{\Delta z}_f}{\Delta z_0}\right)^{-\frac{1}{2}} , \tag{3.45}$$

so that $p(\delta_f^{\rm IGM}) = G(\ln(-\ln T); 0; \sigma_f^{\rm IGM})$. The IGM model flux deficit for filter f is then:

$$\Delta \log F_\nu^{\rm IGM} = -\frac{e^{\delta_f^{\rm IGM}}}{\ln 10} \left[e^{b_1^{\rm IGM} + \tan\theta_1^{\rm IGM}(z_{\rm abs}-z_1)} + e^{b_2^{\rm IGM} + \tan\theta_2^{\rm IGM}(z_{\rm abs}-z_2)} \right] . \tag{3.46}$$

3.5 Afterglow Model Fitting Procedure

3.5.1 Good (and Bad) Modeling Practices

Having constructed our GRB emission, extinction and absorption models, it is important to step back and review good modeling practices in general – and to highlight mistakes that are commonly made, and how to avoid them. We consider the general problem of fitting a model to optical photometric observations of a GRB afterglow. In this case, the data are a set of measured flux densities, obtained from analysis of CCD images taken through a variety of photometric filters, using a variety of cameras on a variety of telescopes, and analyzed by a variety of people using a variety of techniques.

3.5.1.1 Bring the Model to the Data

In all of our modeling efforts, our constant mantra is "Bring the model to the data." In other words, the measured data points, and their associated statistical uncertainties, are "sacrosanct". "Correcting the data", or "bringing the data to the model", requires error propagation, which can seldom be done correctly because of the differences between random and systematic errors, among other pitfalls. As a simple example of bringing the model to the data, consider the problem of accounting for the emission of a GRB host galaxy. A common, but wrong, approach is to wait until the afterglow has faded away, measure the flux density of just the host galaxy using the same filters, and then subtract the measured host galaxy flux density from the afterglow flux density data. This is problematic because there will be uncertainty in the measured host galaxy flux density, consequently *biasing* the subtracted afterglow data either high or low. Some people try to compensate for this by adding additional uncertainty, typically in quadrature, to that of the afterglow measurements. But this is also wrong, since these uncertainties are random and cancel out in the limit of a large number N of afterglow

measurements; the uncertainty introduced by subtracting the measured flux density of the host galaxy is *systematic*, not random.

Although the effect of such a systematic uncertainty might well be small, it would need to be shown as such. However, the most convincing way to do this is simply to model the data correctly in the first place. In this case, the proper approach is to include the host galaxy flux density in each filter as free parameters in the intrinsic emission model itself, and then to find their best-fit values and uncertainties by fitting to *all* of the observed data, including both earlier and later measurements.[8]

As a more subtle example of bringing the model to the data, consider the modeling of extrinsic scatter, or slop, in the afterglow's flux density. Since a CCD camera is an intrinsically linear flux-measuring device, the intrinsic uncertainties in the measurements are expected to be normally distributed in linear F_ν space. However, we treat unmodeled, lower-level, shorter-timescale variability, or slop, in the GRB afterglow flux density to be normally distributed in $\log F_\nu$ space; hence, as discussed in §2.7, we must take care to properly transform the log-normal model distribution into linear F_ν space, in order to compare it to the data and compute the likelihood function. We will review this procedure in some detail in §3.5.2, below.

3.5.1.2 Fit Spectral and Temporal Models Simultaneously

Our observational data sets consist of flux density measurements of GRB afterglows over a range of times and frequencies; they are inherently three-dimensional data sets. A common, but wrong, approach in fitting models to afterglows (and analogous astrophysical phenomena) is to fit using effectively two-dimensional models. For example,

[8]This is correct for ground-based photometric measurements, at least, where the GRB afterglow and the host galaxy are both unresolved. In images from space-based telescopes, it is often possible to separate the afterglow from the host galaxy in photometric measurements. In this case, data from ground-based telescopes and data from space-based telescopes should be placed in separate *model groups*; the host galaxy flux model should be applied when fitting to the former model group, but not when fitting to the latter. All other parameters should be linked between the model groups.

one approach is to fit a light curve model to the data in each photometric filter separately, and obtain a set of fitted light curve normalizations at some common time. Then, one can fit a spectral energy distribution model to these fitted normalizations. This approach is problematic for several reasons. If the fits to the light curves were not done correctly, e.g., if slop was not modeled as a free parameter, or was not modeled at all, the uncertainties in the fitted normalizations could be underestimated, sometimes greatly so (§2.1). On the other hand, if the common normalization time does not coincide with a natural pivot point in the data (§2.6.6), these uncertainties could be greatly overestimated. Then there is the question of how the error bars in the the fitted normalizations are actually computed: Are they cross-sectional (underestimated), marginalized, or projected (overestimated) (§2.6.4)? Are they assumed to be symmetric when they really are not? Can they be well modeled as Gaussians or asymmetric Gaussians in the spectral fitting? Sometimes, there is insufficient data in a single photometric filter to fit a light curve model. In this case, it is common practice to combine data from multiple filters, using an estimated (or iteratively fitted) spectral model to scale them to a common frequency. This is problematic because there will be uncertainties in the estimated (or iteratively fitted) values used to scale the data. These uncertainties are seldom then folded into the temporal fit, but even if they are, they are incorrectly treated as random when, in fact, they are systematic. Rather than having to demonstrate that all of these, and potentially other, problems and potential problems are negligible, it is simpler – and better – to fit to the data directly whenever possible. In this case, the correct approach is to fit a three-dimensional model of flux density as a function of time and frequency simultaneously to the data.[9]

[9]Sometimes, when visualizing the data and the best-fit model, it may be necessary to shift the data to a common time (or frequency) for ease or clarity in plotting, but this should only be done after obtaining the best three-dimensional fit.

3.5.1.3 Integrate Spectral Models Over Filter Response Functions

Another issue arises when data obtained with a photometric filter is taken to have been obtained at a single, effective frequency. This is problematic, because fitted spectral models can then miss, or overemphasize, spectral features that are narrower than the filter bandwidth (i.e., the UV bump due to dust extinction, the damped Lyα absorption line, the Lyα forest and the Lyman limit). Some people try to compensate for this problem with an iterative approach: by first finding a best-fit spectral model, then integrating that best-fit model over each filter's response function to determine new effective frequencies, then re-fitting, integrating and iterating until a stable solution is obtained. While this approach may lead to a valid best-fit model, the problem comes when computing the error bars on best-fit parameter values; this requires evaluation of other, not best-fit, models that might require different iterated effective frequencies. To hold the effective frequencies fixed at the iterated best-fit values disfavors models that favor different effective frequencies, leading to underestimated error bars on the best-fit parameters. This is akin to computing cross-sectional error bars (§2.6.4). The correct, though admittedly computationally expensive, approach is to integrate the spectral flux density model using the filter response functions to obtain a filter-weighted mean flux density at every step, or in the inner loop, of the fitting process, i.e., using these filter-weighted mean model flux densities together with the measured flux densities to compute the likelihood function.

3.5.1.4 Fit Slop as a Free Parameter

To account for shorter-timescale, lower-level unmodeled variability, or slop, in a GRB afterglow, a common practice is to fit using a χ^2 statistic, and to "inflate" the intrinsic error bars of the data (either by adding a constant in quadrature, or by multiplying by some constant factor) until $\chi^2/(\text{degrees of freedom}) \simeq 1$. As we discussed in Chapter 2,

this practice is tantamount to holding the slop parameter fixed at its best-fit value. Any uncertainties in the other fitted model parameters are then cross-sectional with respect to the slop parameter (§2.6.4), and will hence be underestimated. The proper approach, of course, is to fit to slop as a free model parameter.

3.5.1.5 Fit to Calibration Offsets as Model Parameters

Besides unmodeled intrinsic variability, the observational data are also subject to *calibration uncertainty*; in the case of optical photometry using CCD images, calibration uncertainty depends on the accuracy of standard star photometry in the field of view, differences between the combined response functions of the filters and instruments from those used to compile the standard star catalogs, and differences in the data reduction techniques employed by various observers. A common modeling practice is to effectively subsume this calibration uncertainty into the slop – say, by adding it in quadrature to the intrinsic uncertainties of the data points. The problem with this approach is that it treats calibration uncertainties as random errors, which will cancel out in the limit of a large number N of data points, and not as systematic errors, which will propagate independently of N.

The proper approach is to include calibration offsets as model parameters. As we discussed in §3.1, with our genetic algorithm software package *Galapagos*, we can assign different subsets of the data into *calibration groups* – typically, one group per photometric filter per independently prepared data set. Each group is assigned a calibration offset parameter Δ_c in $\log F_\nu$ space, constrained by a zero-mean, log-normal Gaussian prior, $p(\Delta_c) = G(\Delta_c; 0, \sigma_{\log F_{\nu c}})$ whose width $\sigma_{\log F_{\nu c}}$ is a best-guess value for the magnitude of the calibration uncertainty. As we describe below in §3.5.2, when comparing the data in a given calibration group to the model distribution to compute the likelihood function, this offset is added to the model $\log F_\nu$ before transforming the

model distribution into linear F_ν space, effectively shifting the model distribution *en masse* for that entire group. Each calibration offset is thus a fitted model parameter, constrained by a prior.

3.5.1.6 Fit Data Groups and Nested Models with Parameter Linking

As we discussed in §3.1 and §2.6.2, there are other ways one may wish to subdivide the data into subgroups. A common situation involves fitting models with certain parameters that change (possibly discontinuously) in time. For instance, a GRB may experience a sudden rebrightening, due to a discontinuity in the circumburst medium density profile, or due to variable energy injection into the jet (see §4.5). Or, the burst and afterglow may modify the dust in the circumburst medium (or spread laterally and illuminate unmodified populations of dust) as it evolves. In such cases, it may be necessary to divide the data into "time slices", and fit different models to each time slice separately. A common practice is to do exactly that: Divide the data into time slice subgroups, and fit models to each group independently, with all parameters free in each fit. The problem with this approach is that there are some model parameters that we might expect *not* to change with time (but which are nonetheless separate parameters between the time slices), like those describing dust extinction in the Milky Way or absorption due to neutral hydrogen in the IGM. In that case, the best that can be done with this technique is to average the best-fit values of those parameters (or otherwise "merge" their fitted probability distributions) for all the time slices. *Galapagos* allows us to take a more flexible approach: We can divide the data into time slices, and "link" certain parameters across some or all of the slices, while allowing others to vary independently from slice to slice, and fit to all of the time slices simultaneously. The linked parameters are still free to vary to find the best fit, but for any given fit, they will always have the same value in each time slice. The probability distributions of the

fitted parameters can then be computed directly and simultaneously for each linked and unlinked parameter, without any averaging or merging of independent fits.

Parameters may also be *functionally linked*, so that the value of one may be a function of the values of one or more of the others; for instance, in the standard relativistic blast wave GRB emission model, the spectral and temporal indices β and α in different regimes can be expressed as various functions of the electron energy distribution index p and the circumburst density profile index k (see §3.2). Functional linking allows us to explore the relative likelihoods of *nested models*, or sub-parameterizations of a given model (§2.6.2). For example, by first finding the best-fit model with k as a free parameter, the relative likelihoods of the best-fit constant density ($k = 0$) and wind-driven ($k = -2$) circumburst medium models can be quantified.

3.5.2 Model Distribution and the Likelihood Function

The emission, extinction and absorption models described in §§3.2, 3.3 & 3.4 allow us to calculate, at any given time and frequency, the expected model flux density $F_\nu^{\rm mod}(t, \nu; \vartheta_m)$, given a set of M model parameters $\{\vartheta_m\}$. We now must calculate the likelihood of a given model, given a particular set of N filter-integrated flux density measurements $\left\{F_{\nu,n}^{\rm obs}\right\}$, along with any prior constraints on the model parameters themselves. We denote a given flux measurement $F_{\nu,n}^{\rm obs}(t_n, f_n(\nu))$, where $n = \{1, \ldots, N\}$, t_n is the time since the burst trigger at the middle of a given exposure, and $f_n(\nu)$ is the normalized transmission function of the observation's photometric filter, multiplied by the response function of the telescope and instrument, when available. Tabulated response curves are read in at the beginning of the fit for each data point n, and resampled, using linear interpolation, to create normalized filter response curves of N_n points, evenly spaced in linear frequencies $\nu_n = \{\nu_{n,1}, \ldots, \nu_{n,N_n}\}$. In practice, we choose sampling densities that produce approximately constant binwidths $\Delta\nu$ for all filters,

where $\Delta\nu$ is chosen to give $\sim$10 bins across the width of the narrowest spectral feature in the model; in most cases, this is the UV bump, so that the typical bin width is $\Delta\nu \simeq \frac{3}{10}\frac{\gamma}{1+z_{\rm GRB}} \times 10^{14}$ Hz $\approx 2.5 \times 10^{13}$ Hz$/(1+z_{\rm GRB})$. The filter response-weighted mean model flux density is then approximated as:

$$F^{\rm mod}_{\nu,n}(t_n, f_n(\nu); \theta) = \sum_{k=1}^{N_n} F^{\rm mod}_{\nu}(t_n, \nu_{n,k}; \vartheta_m) \times f_n(\nu_{n,k})\,. \tag{3.47}$$

Each flux density measurement $F^{\rm obs}_{\nu,n}$ is assigned to a particular *calibration group*, typically one group per photometric filter per independently prepared data set. Since photometric calibration is typically performed in magnitude space, we assume that all flux measurements within a given calibration group may be offset from their true values by a constant offset $\Delta_{c,n}$ in $\log F$ space. These offsets are parameters in our model, each of which is constrained by a zero-mean log-normal Gaussian prior whose width is given by our best estimate of the calibration uncertainty, $\sigma_{\log F_{\nu_c,n}}$, for that group, i.e., $p(\Delta_{c,n}) = G(\Delta_{c,n}; 0, \sigma_{\log F_{\nu_c,n}})$. The model flux density must then be adjusted to compensate for this offset:

$$F^{\rm mod}_{\nu_c,n} = F^{\rm mod}_{\nu_c,n} \times 10^{\Delta_{c,n}}\,. \tag{3.48}$$

In addition to calibration group offsets, we assume that there is additional, random scatter in the observed flux that is greater than can be attributed to measurement uncertainties alone, due to unmodeled stochastic physical processes in the emission and absorption of light from the GRB afterglow. This additional scatter, or slop, is assumed to be normally distributed in $\log F_\nu$ space, and is characterized as a Gaussian with width $\sigma_{\log F_\nu}$. While, in principle, this slop could be different for different data groups, in practice we assume that $\sigma_{\log F_\nu}$ is the same for all measurements in the UV-optical-IR region of the spectrum, and we characterize it with a single model parameter.[10] In this

[10]However, when we include X-ray flux measurements in our data set, we employ a separate slop parameter $\sigma^{\rm X}_{\log F}$ (see §4.3.2). If we include radio observations, there may be an additional component

sense, we are fitting not a model curve, but a model distribution to our measured flux densities. The model probability density is normally distributed about the model curve in $\log F_\nu$ space:

$$p^{\rm mod}(\log F_\nu; \log F^{\rm mod}_{\nu_c,n}, \sigma_{\log F_\nu}) = \frac{1}{\sqrt{2\pi}\sigma_{\log F_\nu}} e^{\frac{\left(\log F_\nu - \log F^{\rm mod}_{\nu_c,n}\right)^2}{2\sigma^2_{\log F_\nu}}} . \tag{3.49}$$

Since we are comparing the model distribution to flux densities in linear space, it is necessary to transform the probability distribution of the model into that space. As we showed in §2.7, the probability density of the model distribution in linear F_ν space is obtained by equating the differential probabilities:

$$\begin{aligned} p^{\rm mod}(F_\nu; F^{\rm mod}_{\nu_c,n}, \sigma_{\log F_\nu}) dF_\nu &= p(\log F_\nu; \log F^{\rm mod}_{\nu_c,n}, \sigma_{\log F_\nu}) d\log F_\nu \\ p^{\rm mod}(F_\nu; F^{\rm mod}_{\nu_c,n}, \sigma_{\log F_\nu}) &= \frac{1}{F_\nu \ln 10} p(\log F_\nu; \log F^{\rm mod}_{\nu_c,n}, \sigma_{\log F_\nu}) . \end{aligned} \tag{3.50}$$

It can be shown analytically that $p^{\rm mod}$ peaks at a value:

$$F^{\rm max}_{\nu_c,n} = 10^{\left(\log F^{\rm mod}_{\nu_c,n} - \sigma^2_{\log F_\nu} \ln 10\right)} . \tag{3.51}$$

As we discussed in §2.7, it is often convenient to approximate asymmetric probability distributions as skew-normal, or asymmetric, Gaussian distributions. As a reminder, we define the asymmetric Gaussian function:

$$G_A(x; x_0, \sigma_{x\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{x+} + \sigma_{x-})} \times \begin{cases} e^{-\frac{1}{2}\frac{(x-x_0)^2}{\sigma^2_{x+}}} & \text{if } x \geq x_0 \\ e^{-\frac{1}{2}\frac{(x-x_0)^2}{\sigma^2_{x-}}} & \text{if } x < x_0 \end{cases} . \tag{3.52}$$

If $\sigma_{\log F_\nu} \lesssim 0.1$, $p^{\rm mod}(F_\nu)$ can be reasonably well approximated as $G_A(F_\nu; F^{\rm max}_{\nu_c,n}, \sigma_{F_\nu\pm})$,

of slop due to scintillation in the ISM, which is expected to be normally distributed in linear F_ν space; this would have to be modeled separately as well.

where $\sigma_{F_\nu \pm}$ are approximated by the empirical formulas:

$$\begin{aligned}
\sigma_{F_\nu +} &\approx F^{\max}_{\nu_c,n} \left(2.3029\sigma_{\log F_\nu} + 2.6293\sigma^2_{\log F_\nu} - 3.6945\sigma^3_{\log F_\nu}\right) \\
\sigma_{F_\nu -} &\approx F^{\max}_{\nu_c,n} \left(2.3027\sigma_{\log F_\nu} - 2.6544\sigma^2_{\log F_\nu} - 4.0699\sigma^3_{\log F_\nu}\right) .
\end{aligned} \tag{3.53}$$

As we described in detail in Chapter 2, to find the best-fit model distribution to the observed data, we must define a likelihood function that quantifies the joint probability of a given set of measurements, with their intrinsic measurement uncertainties, and the model distribution described by the set of parameters $\{\vartheta_m\}$, which includes one or more slop parameters σ. We approximate the intrinsic probability distribution of a given flux density measurement $F_{\nu,n}$ as a (possibly asymmetric) Gaussian:

$$p^{\rm int}_{\nu,n}(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}\pm}) = G_A(F_\nu; F_{\nu,n}, \sigma_{F_{\nu,n}\pm}) . \tag{3.54}$$

For optical and NIR photometric observations, since a CCD camera is an intrinsically linear flux-measuring device, the flux density measurements we generate ourselves have symmetric error bars $\sigma_{F_{\nu,n}+} = \sigma_{F_{\nu,n}-} = \sigma_{F_{\nu,n}}$. Published data from other sources, however, tend to quote symmetric error bars in $\log F_\nu$ or magnitude space. In those cases, we transform the data points into linear flux space in the same manner as outlined above for the log-normal model distribution (assuming the quoted error bars are $\lesssim 0.1$ in $\log F_\nu$), with asymmetric error bars $\sigma_{F_{\nu,n}\pm}$. If a published measurement is quoted as a 3σ upper limit, we assign that data point a linear flux density $F_{\nu,n} = 0$, and assign it symmetric 1σ error bars $\sigma_{F_{\nu,n}}$ such that:

$$\int_{-\infty}^{F^{3\sigma}_{\nu,n}} G(F_\nu; 0, \sigma_{F_{\nu,n}}) dF_\nu = 0.99730 , \tag{3.55}$$

as we discussed in §2.7.3. If 2- or 1σ upper limits, $F^{2\sigma}_{\nu,n}$ or $F^{1\sigma}_{\nu,n}$, are given, the values

of $\sigma_{F_{\nu,n}}$ are those for which the integral in Equation 3.55 equals 0.9545 and 0.6827, respectively. By these definitions, $\sigma_{F_{\nu,n}} = F^{3\sigma}_{\nu,n}/2.7822$, $\sigma_{F_{\nu,n}} = F^{2\sigma}_{\nu,n}/1.6901$ or $\sigma_{F_{\nu,n}} = F^{1\sigma}_{\nu,n}/0.4752$.

The joint probability of the one-dimensional model distribution and a given flux density measurement is:

$$p_n \propto \int_{-\infty}^{\infty} G_A(F_\nu; F^{\max}_{\nu_c,n}, \sigma_{F_\nu \pm}) G_A(F_\nu; F_{\nu,n}; \sigma_{F_{\nu,n}\pm}) dF_\nu \,. \tag{3.56}$$

If both distributions were symmetric Gaussians, with widths σ_{F_ν} and $\sigma_{F_{\nu,n}}$, respectively, this joint probability takes a very simple form:

$$p_n \propto G(F_{\nu,n}; F^{\max}_{\nu_c,n}, \Sigma_n) \,, \tag{3.57}$$

where Σ_n is the quadrature sum of the slop and the intrinsic uncertainty:

$$\Sigma_n \equiv \sqrt{\sigma^2_{F_\nu} + \sigma^2_{F_{\nu,n}}} \,. \tag{3.58}$$

However, when one or both of the distributions is approximated as an asymmetric Gaussian, the result of the probability integral (equation 3.56) is not, in general, of the form of an asymmetric Gaussian. We have explored such integrals extensively (see §2.7 and Appendix A), and have found that they can be reasonably well approximated as asymmetric Gaussians, provided the peak is allowed to shift by an amount $\delta_{F_{\nu,n}}$. An approximate empirical formulation for $\delta_{F_{\nu,n}}$ is presented in Appendix A. The joint probability can then be approximated as:

$$p_n \propto G_A\left(F_{\nu,n} + \delta_{F_{\nu,n}}; F^{\max}_{\nu_c,n}, \Sigma_{n\pm}\right) \,, \tag{3.59}$$

where

$$\begin{aligned} \Sigma_{n+} &\equiv \sqrt{\sigma_{F_\nu -}^2 + \sigma_{F_{\nu,n}+}^2} \\ \Sigma_{n-} &\equiv \sqrt{\sigma_{F_\nu +}^2 + \sigma_{F_{\nu,n}-}^2}\,. \end{aligned} \tag{3.60}$$

Note the reversal of signs; $\sigma_{F_\nu \mp}$ is added in quadrature to $\sigma_{F_{\nu,n}\pm}$.

The total likelihood of the measurements, given the model, is the product of the individual joint probabilities:

$$\mathcal{L} = \prod_{n=1}^{N} p_n\,. \tag{3.61}$$

However, some of our model parameters $\{\vartheta_m\}$ themselves have prior constraints on their values. Typically, these are symmetric or asymmetric Gaussian probability distributions about the most likely parameter value. In some cases, one parameter's prior width or peak may be a function of one or more other parameters. The details of the priors that we use in our GRB afterglow extinction and absorption models are discussed in detail in §§3.3 & 3.4. Including the priors, the posterior probability of the model, given the data, is:

$$\mathcal{P} \propto \mathcal{L} \times \prod_{m=1}^{M} p_{\vartheta_m}\,. \tag{3.62}$$

Finally, we arrive at the best-fit set of model parameters for a given data set by minimizing $-2\ln\mathcal{P}$:

$$-2\ln\mathcal{P} = \sum_{n=1}^{N} -2\ln\mathcal{L} + \sum_{m=1}^{M} -2\ln p_{\vartheta_m} + \text{Constant}\,. \tag{3.63}$$

This quantity has the form of a modified, χ^2-like statistic (it is not a "true" χ^2, in that it is offset by an arbitrary constant, contains prior information and is a function of one

or more free slop parameters). We find the best-fit model by minimizing:

$$\sum_{m=1}^{M} \frac{(\vartheta_m - \hat{\vartheta_m})^2}{\sigma_{\vartheta,m}^2} + \sum_{i=1}^{N} 2\ln(\Sigma_{n+} + \Sigma_{n-}) + \begin{cases} \frac{(F_{\nu,n}+\delta_{F_{\nu,n}}-F_{\nu_c,n}^{\max})^2}{\Sigma_{n-}^2} & \text{if } F_{\nu,n} + \delta_{F_{\nu,n}} \geq F_{\nu_c,n}^{\max}\,, \\ \frac{(F_{\nu,n}+\delta_{F_{\nu,n}}-F_{\nu_c,n}^{\max})^2}{\Sigma_{n+}^2} & \text{if } F_{\nu,n} + \delta_{F_{\nu,n}} < F_{\nu_c,n}^{\max}\,, \end{cases} \tag{3.64}$$

where the first summation is over all parameters with Gaussian priors $p_{\vartheta_m} = G(\vartheta_m; \hat{\vartheta_m}, \sigma_{\vartheta,m})$, where $\hat{\vartheta_m}$ is the best-fit value of parameter ϑ_m, and $\sigma_{\vartheta,m}$ is its fitted uncertainty. (Note that in some cases, a parameter's prior probability distribution may also be approximated as an asymmetric Gaussian.) The term $2\ln(\Sigma_{n+}+\Sigma_{n-})$ in the second summation is necessary, since $\Sigma_{n\pm}$ is a function of the free slop parameter(s) $\sigma_{\log F_\nu}$. We do not include a term $2\ln\sigma_{\vartheta,m}$ in the prior summation. In some cases, $\sigma_{\vartheta,m}$ is a pre-determined constant. In other cases, $\sigma_{\vartheta,m}$ may itself be a model parameter, with its own prior constraint. We found that in this latter case, including the term $2\ln\sigma_{\vartheta,m}$ biases the parameter $\sigma_{\vartheta,m}$ towards very small values as χ^2 is minimized.

Chapter 4

Fit the First: Exercising the Model on GRB 090313

"For the Snark's a peculiar creature, that won't
Be caught in a commonplace way.
Do all that you know, and try all that you don't:
Not a chance must be wasted to-day!"
— Lewis Carroll, *The Hunting of the Snark*, Fit the Fourth

4.1 Introduction

In this chapter, we exercise the GRB afterglow model and fitting procedure described in Chapter 3 by fitting to NIR, optical and X-ray observations of GRB 090313. With a spectroscopic redshift of $z_{\rm GRB} = 3.375$ (Chornock et al., 2009b), this burst is ideal for testing our UV source-frame dust extinction model (§3.3.3), which affects NIR and especially optical emission in the observer's frame; e.g., the 2175 Å UV bump feature (§3.3.4.3), if present, would lie in the I and z' filters. Furthermore, at this redshift we expect a significant (≈60%) amount of absorption due to the Lyα forest, which overlaps the B, V and g' photometric filters (§3.4.2). Any source-frame damped Lyα absorption line would lie in the middle of the V and g' filters, and the Lyman limit lies in the blue wing of the B filter (see Figure 4.1). The majority of the $BVRI$ photometric

observations were conducted by telescopes in UNC's PROMPT (Reichart et al., 2005) array in Cerro Tololo, Chile, which is part of UNC's Skynet Robotic Telescope Network.

4.2 Observations

The afterglow of GRB 090313 exhibits an interesting, somewhat unusual "slow-riser" light curve, which peaks at $\log t_{12} \approx -2$ days (≈ 14 min) after the initial GRB trigger (see Figure 4.5). As discussed in §3.2, rising light curves may be due to: the blast wave reaching the deceleration radius; our relativistic viewing cone broadening to overlap a potentially brighter central structure of an off-axis jet as it decelerates; the blast wave moving through a region of increasing density; and energy injection by a "refreshed shock" due to later-time ejecta catching up with the decelerating blast wave (e.g., Perley et al., 2009).

GRB 090313 triggered the *Swift* BAT detector at 09:06:28 UT on 13 March 2009 (Mao et al., 2009). The BAT position was RA = 13h13m35s, Dec = $+08°06'23''$ (J2000), with an uncertainty of $3'$. It was later localized by the Katzman Automatic Imaging Telescope (KAIT) at Lick Observatory to RA = 13h13m36.21s, Dec = $+08°05'49.8''$ (J2000) (Chornock et al., 2009a), or galactic coordinates $l = 319.^{\circ}3741, b = 70.^{\circ}2582$. Interpolation of the all-sky IR dust emission maps of Schlegel et al. (1998) provides a local dust extinction color excess $E(B-V)_{\rm MW} = 0.0279$ mag along this line of sight through the Milky Way. Spectrographic observations of the afterglow conducted 19 hours after the burst trigger with Gemini South (Chornock et al., 2009b) determined that the burst occurred at a redshift $z_{\rm GRB} = 3.375$, ruling out an association with a relatively bright SDSS galaxy at $z = 0.0235$ in the field. Due to moon proximity constraints, the *Swift* XRT instrument was not able to begin taking measurements of this afterglow until 7.4 hours after the burst trigger (Mao & Margutti, 2009).

We observed the afterglow of GRB 090313 using three 16-inch telescopes of the

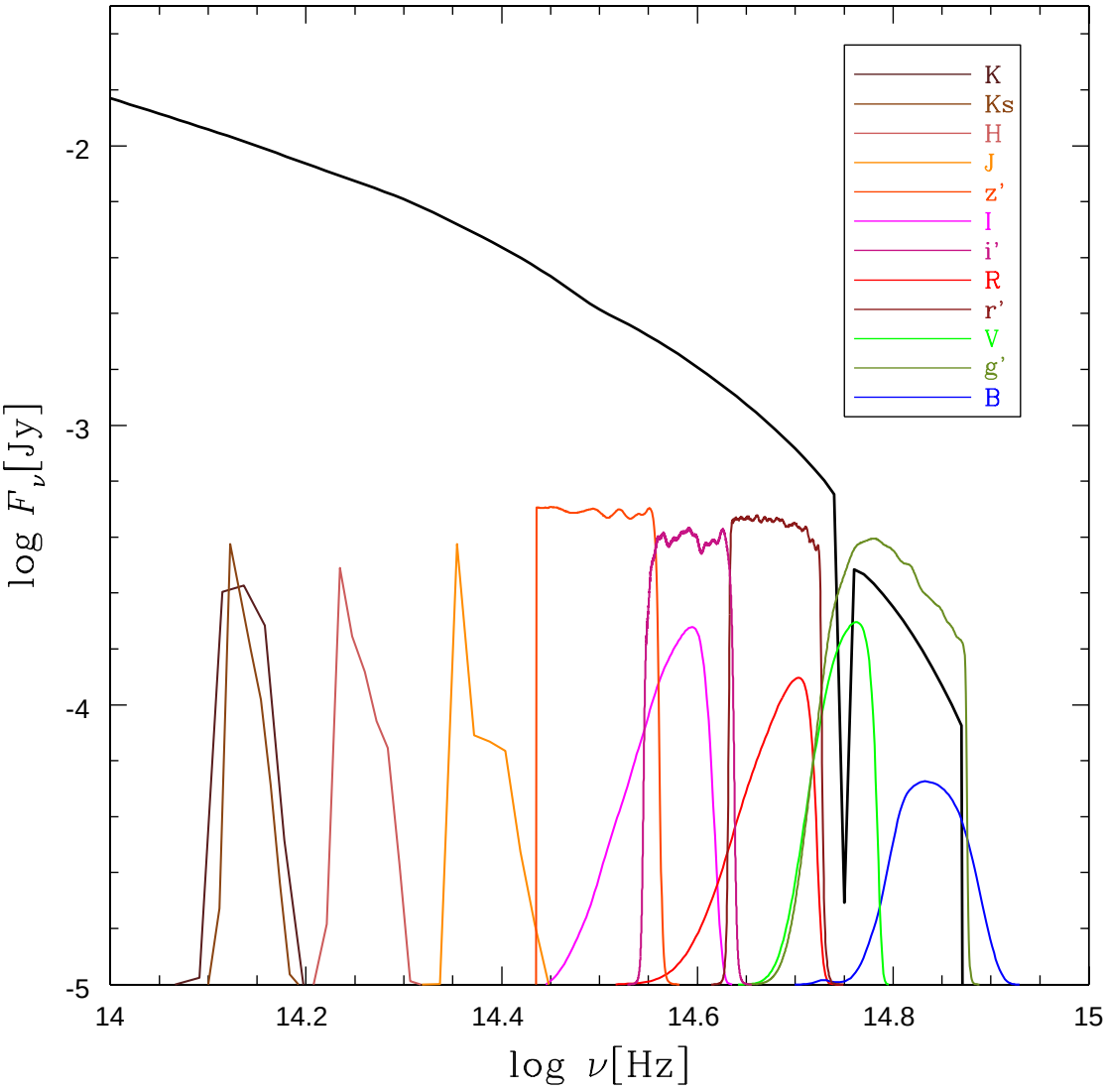

Figure 4.1 Fitted constant-density ($k = 0$) model NIR–X-ray spectrum (black curve) for GRB 090313 at the light curve peak, $logt_{12} \approx -2$ days. NIR and optical photometric filter transmission functions are plotted along the bottom axis. $BVRI$ transmission functions have been convolved with the response functions of the relevant PROMPT CCD cameras. High-density functions are used at the high frequencies. See §4.4 for a full discussion of this fit.

PROMPT array at Cerro Tololo, Chile, beginning 133 seconds after the *Swift* trigger. We took exposures ranging between 10 and 80 seconds in duration through Johnson-Cousins B, V, R and I filters for the next 42 minutes until dawn. We resumed observations the following night, obtaining $BVRI$ detections and upper limits from stacked images between 17.3 and 24.2 hours after the burst trigger. The 16-inch telescope at Dolomites Astronomical Observatory at Madonna di Campiglio, Italy, which is part of the Skynet Robotic Telescope Network, obtained an additional detection in the R band at 13.8 hours. All image calibration, stacking and photometry was performed with Skynet's web-based data reduction pipeline and image analysis software, *Afterglow* (Ivarsen, in preparation; Haislip, in preparation; `http://skynet.unc.edu/afterglow/`). *Afterglow* reproduces IRAF's `DAOPHOT` package for photometry, and computes flux densities using published zero-point values for each photometric filter. All photometric measurements generated by *Afterglow* are assigned symmetric Gaussian 1σ uncertainties in linear flux density space; flux density measurements for which the integrated probability distribution at $F_\nu > 0$ is less than 0.9973 (§2.7.3) are plotted as 3σ upper limits in $\log F_\nu$ space (see Figure 4.5). The data are divided into calibration groups: one group per filter, per telescope. *Afterglow* computes estimated 1σ calibration uncertainties in $\log F_\nu$ space for each group; these provide the width of the zero-mean Gaussian prior for each group's calibration offset model parameter $\Delta_{c,n}$ (§3.5.2, Equation 3.48).

To broaden the spectral and temporal coverage of our data set, we include a number of observations published in GCN Circulars in the days immediately following the burst. These include: three observations from the 1.3-meter PAIRITEL telescope at Mt. Hopkins, Arizona at $t = 0.941$, 2.302 and 3.512 days in each of the J, H and K_S NIR filters (Morgan et al., 2009); three observations from the GROND instrument on the 1.2 m telescope at La Silla, Chile at $t = 10$ min in the J, H, and K filters (Updike

et al., 2009); $g'r'i'z'$ observations at $t = 20.07$ hr from the 8.1-meter Gemini South telescope at Cerro Pachon, Chile (Perley, 2009); four J-band observations between $t = 0.805$ and 3.834 days from the 1.3 m CTIO telescope in Chile (Cobb, 2009); and one K_S-band observation at $t = 19.17$ hr from the 2.5-m Nordic Optical Telescope (NOT) at La Palma, Canary Islands (de Ugarte Postigo et al., 2009). For these observations, which were calibrated using either 2MASS or SDSS reference stars, we assume a fiducial calibration uncertainty of 0.02 in $\log F_\nu$, which provides the 1σ width of the zero-mean Gaussian priors on the calibration offset parameters $\Delta_{c,n}$ (§3.5.2, Equation 3.48) for each calibration group (one group per filter per telescope). The published magnitudes, with their quoted symmetric Gaussian uncertainties, and any 3σ upper limits were transformed into linear flux density space using the techniques described in §2.7.3. We also include X-ray data collected by the *Swift* XRT instrument (Evans et al., 2009) from 0.31 to 9.28 days after the burst trigger. We describe these data, and the details of our X-ray flux model fitting procedure in §4.3.2.

4.3 Emission Model

4.3.1 Parameterization of the Blast Wave Emission Model

We model the intrinsic emission of a GRB afterglow using the standard blast wave model for synchrotron emission from shocks in a decelerating relativistic jet (§3.2, and references therein). The light curve and spectrum are described by a series of smoothly broken power laws in time and frequency. The simplest emission model is a single power law with no breaks, where the emitted flux density at time t and frequency ν in the observer's frame is:

$$F_\nu(t) = F_0 \left(\frac{t}{t_0}\right)^\alpha \left(\frac{\nu}{\nu_0}\right)^\beta . \tag{4.1}$$

where t_0 and ν_0 are a fixed reference time and frequency, typically set equal to the mean log time and log frequency of observations in the data set. The reference flux density F_0 is a free parameter in the model. Note that in our notation, positive indices describe rising light curves and spectra, while negative indices describe falling ones.

The next simplest model is a fading light curve with a cooling break but no jet break. The temporal and spectral indices can be expressed as functions of the electron energy index $p < -2$ in the jet, and the circumburst medium radial density index k, where $N(E) \propto E^p$ and $n(r) \propto r^k$. Radial density index $k = 0$ corresponds to a constant-density (ISM) medium (e.g., Sari et al., 1998), and $k = -2$ corresponds to an idealized stellar wind model (e.g., Chevalier & Li, 1999). The spectral and temporal indices as a function of p and k have been derived by Starling et al. (2008).

$$\begin{aligned} \alpha_{a2} &= \tfrac{12(p+1)+k(3p+5)}{4(4+k)}, \quad \beta_{a2} = \tfrac{p+1}{2} \quad (\nu < \nu_c(t)) \\ \alpha_{b2} &= \tfrac{3p+2}{4}, \quad \beta_{b2} = \tfrac{p}{2} \quad (\nu > \nu_c(t)) \ . \end{aligned} \tag{4.2}$$

where subscripts a and b indicate, respectively, the indices at frequencies below and above the cooling break frequency $\nu_c(t)$ at any given time. In our model, the electron energy index p is a free parameter, while the circumburst density index k can either be free, or be set fixed at certain canonical values to test the relative fitness of different circumburst density models. As before, we select a fixed reference time and frequency t_0 and ν_0. The flux F_0 at this time and frequency is a free parameter in our model, as is the frequency ν_{c0} of the cooling break at the reference time t_0. By equating the flux due to the two power laws before and after the cooling break (Equation 4.1), we can solve for the frequency of the cooling break at any given time. We find that the time dependence of the cooling break frequency for arbitrary circumburst density index k

can be expressed as:

$$\nu_c(t) = \nu_{c0} \left(\frac{t}{t_0} \right)^{f(k)} , \tag{4.3}$$

where

$$f(k) = \frac{1}{1 + k/4} - \frac{3}{2} . \tag{4.4}$$

Note that, for the uniform density ISM case ($k = 0$), the cooling break moves from high to low frequencies as $\nu_c(t) \propto t^{-0.5}$, while for the stellar wind case ($k = -2$), it moves from low to high frequencies as $\nu_c(t) \propto t^{0.5}$, as has been reported in previous publications (e.g., Sari et al. (1998), Chevalier & Li (1999)).

At any given time, it is possible to compute the (unsmoothed) flux density at the cooling break, $F_c(t) = F_{\nu_c(t)}$ from the reference flux density F_0 and the spectral and temporal indices of Equations 4.2. The model spectrum at that time can then be expressed as a smoothly broken power law (§3.2.2):

$$F_\nu(t) = F_c(t) \left[\left(\frac{\nu}{\nu_c(t)} \right)^{-\beta_a s_c} + \left(\frac{\nu}{\nu_c(t)} \right)^{-\beta_b s_c} \right]^{-1/s_c} , \tag{4.5}$$

where β_a and β_b, the spectral indices below and above the cooling break frequency, respectively, are given by Equation 4.2, and where the smoothing parameter is an semi-empirical function of the electron energy index p and the circumburst medium density index k (Granot & Sari, 2002). For the constant-density ($k = 0$) and wind-blown ($k = -2$) cases:

$$\begin{aligned} s_c &\approx 1.15 + 0.06p \text{ if } k = 0 \\ s_c &\approx 0.80 + 0.03p \text{ if } k = -2 . \end{aligned} \tag{4.6}$$

(For $k \neq 0$, -2 we simply interpolate or extrapolate from these values.)

We now consider the case of a light curve that rises to a peak at early times before

fading, as is the case with GRB 090313. We indicate the temporal and spectral indices prior to the peak and after the peak by subscripts 1 and 2, respectively. In our model, we choose to set $t_0 = t_{12}$, i.e., the time of the light curve peak, and allow t_{12} to vary as a free model parameter. In this case, F_0 corresponds to the unsmoothed peak flux at the (fixed) reference frequency ν_0. Given the sparseness of our data set in the pre-peak regime, we simply assume that the temporal indices before the light curve peak, both below and above the cooling break, are offset by a constant $\Delta\alpha_1 > -\alpha_2$ from their corresponding post-peak indices, and that the spectral index at a given observed frequency does not change across the peak:

$$\begin{aligned} \alpha_{a1} &= \alpha_{a2} + \Delta\alpha_1, \quad \beta_{a1} = \beta_{a2} = \tfrac{p+1}{2} \\ \alpha_{b1} &= \alpha_{b2} + \Delta\alpha_1, \quad \beta_{b1} = \beta_{b2} = \tfrac{p}{2} \quad , \end{aligned} \tag{4.7}$$

where the post-peak indices are given by Equations 4.2. Since the spectral indices are unchanged, and the pre-peak temporal indices at frequencies both below and above the cooling break differ from the post-peak indices by the same constant, the cooling break frequency as a function of time is the same in both the pre-peak and post-peak regimes (Equation 4.3).

We compute the pre-peak light curve and spectrum independently of the post-peak model (Equation 4.5), as if the light curve were rising throughout the entire duration of our observations. Again, at any given time, it is possible to compute the flux at the cooling break frequency, $F_{c1}(t)$, using the reference flux F_0 and the pre-peak spectral and temporal indices. The pre-peak spectral flux density at a given time then has the same form as Equation 4.5:

$$F_1(t) = F_{c1}(t) \left[\left(\frac{\nu}{\nu_c(t)} \right)^{-\beta_{a1} s_c} + \left(\frac{\nu}{\nu_c(t)} \right)^{-\beta_{b1} s_c} \right]^{-1/s_c} . \tag{4.8}$$

The model flux density is then the smoothed weighted average of the pre- and post-peak flux densities:

$$F_{12}(t) = \left[F_1(t)^{-s_{12}} + F_2(t)^{-s_{12}}\right]^{-1/s_{12}} , \tag{4.9}$$

where $F_1(t)$ is given by Equation 4.8, $F_2(t)$ is given by Equation 4.5, restated here with subscript 2s:

$$F_2(t) = F_{c2}(t)\left[\left(\frac{\nu}{\nu_c(t)}\right)^{-\beta_{a2}s_c} + \left(\frac{\nu}{\nu_c(t)}\right)^{-\beta_{b2}s_c}\right]^{-1/s_c} , \tag{4.10}$$

and where s_{12} is a free, empirical parameter describing the smoothness of the light curve at the peak. Note that a smoothing parameter value $s_{12} \gg 1$ corresponds to a sharp peak in the light curve, while $0 < s_{12} \ll 1$ corresponds to a very gradual turnover.

When the forward edge of the GRB blast wave has swept up a mass of circumburst material roughly equal to the mass of the jet ejecta, it undergoes a sudden deceleration, called the jet break, after which lateral expansion dominates over forward motion (e.g Sari et al., 1999, §3.2.1). The post-jet break temporal and spectral indices are:

$$\begin{aligned} \alpha_{a3} &= p, \quad \beta_{a3} = \tfrac{p+1}{2} \\ \alpha_{b3} &= p, \quad \beta_{b3} = \tfrac{p}{2} \end{aligned} , \tag{4.11}$$

where, as before, subscripts a and b denote indices at frequencies below and above the cooling break, and the subscript 3 is used to indicate the post-jet break temporal regime. Note that in this regime, the temporal indices do not change across the cooling break, and so the cooling break frequency ceases to change with time, i.e., $\nu_{c3}(t) = \nu_{c3}(t_{23}) =$ constant. We proceed to calculate the post-jet break flux $F_3(t)$ in exactly the same manner as the pre-peak flux, with the caveat that the cooling break frequency is a constant, so that the cooling break flux $F_{c3}(t)$ changes only as a function of the

post-jet break temporal index:

$$F_3(t) = F_{c3}(t) \left[\left(\frac{\nu}{\nu_c(t_{23})} \right)^{-\beta_{a3} s_c} + \left(\frac{\nu}{\nu_c(t_{23})} \right)^{-\beta_{b3} s_c} \right]^{-1/s_c} . \tag{4.12}$$

We can compute a smoothed weighted average of the pre- and post-jet break flux densities, analogous to Equation 4.9:

$$F_{23}(t) = \left[F_2(t)^{-s_{23}} + F_3(t)^{-s_{23}} \right]^{-1/s_{23}} , \tag{4.13}$$

where s_{23} is a free, empirical model parameter describing the smoothness of the light curve at the jet break.

The model flux, including both the pre-peak rise and the jet break, can be computed in two ways:

$$F_{12,3}(t) = \left[F_{12}(t)^{-s_{23}} + F_3(t)^{-s_{23}} \right]^{-1/s_{23}} , \tag{4.14}$$

or

$$F_{1,23}(t) = \left[F_1(t)^{-s_{12}} + F_{23}(t)^{-s_{12}} \right]^{-1/s_{12}} . \tag{4.15}$$

Note that, in general, $F_{12,3} \neq F_{1,23}$, though in cases where the smoothing parameter $s \gtrsim 1$, or in cases where $t_{23} \gg t_{12}$, the two combinations are effectively equivalent. However, since there is no clear reason to chose one combination over the other, we take the average of the two, in $\log F$ space, to obtain our model flux density:

$$\log F_\nu^{\rm mod}(t) = \frac{1}{2} \left[\log F_{12,3}(t) + \log F_{1,23}(t) \right] . \tag{4.16}$$

4.3.2 Fitting to X-ray Flux Data

When available, we include any published X-ray flux measurements, post-prompt emission, in our data set (in the case of GRB 090313, from the *Swift* XRT instrument).

The UK *Swift* Science Data Centre publishes: 1) light curves (flux vs. time), corrected for the instrument response function and integrated across a 0.3–10 keV band (`http://www.swift.ac.uk/xrt_curves/00346386/`); and 2) time-averaged spectra, corrected for the instrument response function, both observed and corrected for the modeled absorption due to gas (`http://www.swift.ac.uk/xrt_spectra/00346386/`). From the integrated observed time-averaged spectrum, and the integrated absorption-corrected time-averaged spectrum, we obtained an absorbed-to-unabsorbed flux ratio of 1.375, which we use to correct their light curves for absorption. Finally, in their modeling of the observed, time-averaged spectrum, they measure both an effective unabsorbed spectral index $\overline{\beta}_X^{\rm eff} = -1.14^{+0.14}_{-0.12}$ and an effective source-frame column density $\overline{N}_{H_X}^{\rm eff} = 2.99^{+0.77}_{-0.71} \times 10^{22}\,{\rm cm}^{-2}$ across the 0.3–10 keV band.

The instrument response- and absorption-corrected XRT data are quoted as fluxes (in erg s^{-1} cm^{-2}) integrated over the 0.3–10 keV band, rather than as flux densities (in Jy). We therefore must integrate our model spectral flux density across the 0.3–10 keV band in order to compare it to the data. Since our smoothly broken power-law emission model is not analytically integrable, we approximate the spectrum in the 0.3–10 keV band as a single power law, with an effective spectral index $\beta_X^{\rm eff}$ at time t given by evaluating the emission model at $\nu_{X1} = 0.3\,{\rm keV} = 7.254 \times 10^{16}$ Hz and $\nu_{X2} = 10\,{\rm keV} = 2.418 \times 10^{18}$ Hz:

$$\beta_X^{\rm eff}(t) \equiv \frac{\log F_{\nu_{X2}}^{\rm mod}(t) - \log F_{\nu_{X1}}^{\rm mod}(t)}{\log \nu_{X2} - \log \nu_{X1}}\,. \tag{4.17}$$

We also define an X-ray reference frequency $\nu_X = 4.19 \times 10^{17}$ Hz that is at the mean log frequency of the $0.3 - 10$ keV band. The effective flux density at this frequency is:

$$\log F_{\nu_X}^{\rm eff}(t) = \log F_{\nu_{X1}}^{\rm mod}(t) + \beta_X^{\rm eff}(t)(\log \nu_X - \log \nu_{X1})\,. \tag{4.18}$$

Consequently, the integrated X-ray model flux is then approximated as:

$$F_X^{\rm mod}(t) = \left(\frac{F_{\nu_X}^{\rm eff}(t)}{1+\beta_X^{\rm eff}(t)}\right)\left(\frac{1}{\nu_X}\right)^{\beta_X^{\rm eff}(t)}\left(\nu_{X2}^{1+\beta_X^{\rm eff}(t)} - \nu_{X1}^{1+\beta_X^{\rm eff}(t)}\right) \times 10^{-23}\ {\rm erg\ s^{-1}\ cm^{-2}}\,. \tag{4.19}$$

Finally, we make use of Swift's effective unabsorbed spectral index $\overline{\beta}_X^{\rm eff} = -1.14^{+0.14}_{-0.12}$ in the form of an asymmetric Gaussian prior, $p(\overline{\beta}_X^{\rm eff}) = G_A(\overline{\beta}_X^{\rm eff}; -1.14, +0.14, -0.12)$, where we approximate the time-averaged $\overline{\beta}_X^{\rm eff}$ with the value of $\beta_X^{\rm eff}$ from Equation 4.17 at the flux-weighted mean observing time $t_X = 1.2$ days of the XRT data, i.e., $\overline{\beta}_X^{\rm eff} \simeq \beta_X^{\rm eff}(t_X)$. Since all spectral breaks appear to be at frequencies that are well below the X-ray region of the spectrum during the XRT observation (see §4.4), this is a reasonable approximation. We compute the model distribution and joint probability in the X-ray regime using an independent slop parameter $\sigma_{\log F}^X$.

4.4 Model Fitting Results

By inspection, it is clear from the *Swift* XRT light curve (Figures 4.5 & 4.6) that there is a break at $t \approx 1$ day. The pre- and post-break temporal indices are $\alpha \approx -1.2$ and $\alpha \approx -2.2$, respectively; a change of $\Delta\alpha \approx -1$ is expected across a jet break for electron indices $p \approx -2$, assuming that the cooling break frequency is below the XRT band (Equations 4.2 & 4.11); we therefore include a jet break in our spectral and temporal emission model. The mean effective spectral index of the XRT data is $\overline{\beta}_X^{\rm eff} = -1.14^{+0.14}_{-0.12}$. If the cooling break is below the X-ray band throughout the duration of the observations, the pre- and post-jet break spectral index is expected to be $\beta_{b2} = \beta_{b3} = p/2$, while the post-jet break temporal index is (always) expected to be $\alpha_{3b} = p$ (Equations 4.2 & 4.11). The fact that the observed post-jet break X-ray temporal index of -2.2 is approximately twice the mean effective X-ray spectral index of -1.14 strongly suggests that the cooling break is below the X-ray band throughout

the time range of these data. Furthermore, the observed flux density at the time of the jet break in the K_S band, which is minimally affected by extinction, is consistent with what would be predicted by extrapolating from the observed X-ray flux density using this mean effective spectral index (Figure 4.2). This suggests that, at the time of the jet break, the cooling break is below the optical/NIR region of the spectrum, in which case both the optical/NIR and X-ray spectral indices are expected to be $\beta_{b2} = \beta_{b3} = p/2$ (Equations 4.2 & 4.11).

In the fits that we present below, it is not possible to establish with any degree of certainty whether the best-fit optical/NIR intrinsic model spectrum, corrected for extinction and absorption, is consistent with a constant-density ($k = 0$) or stellar wind ($k = -2$) circumburst density profile. In the former case, the best-fit model places a cooling break in the optical at the light curve peak, which then moves redward past the NIR by the time of the jet break. In the latter case, the cooling break remains below the NIR throughout the time range of our optical/NIR data. But in either case, the cooling break is below the optical/NIR (and certainly below the X-ray band) at the time of the jet break, as the data suggest.

We first attempted to fit a single emission, absorption and extinction model, with one set of model parameters, to the entire time range of the data set. However, there appears to be a rebrightening of the afterglow between the first and second nights that cannot be accounted for with a single set of model parameters, or with acceptable calibration group offsets. We therefore divided the data set into two "time slices" (§3.5.1.6): The first includes all of the PROMPT data taken on the first night, plus any NIR data taken up to $t = 0.146$ days; the second includes the PROMPT and Dolomites data taken on the second night, any remaining NIR data, the Gemini South $g'r'i'z'$ observations, and all of the *Swift* XRT X-ray data ($t = 0.312 - 9.281$ days). The burst redshift $z_{\rm GRB} = 3.375$, the parameters describing Lyα absorption in the IGM,

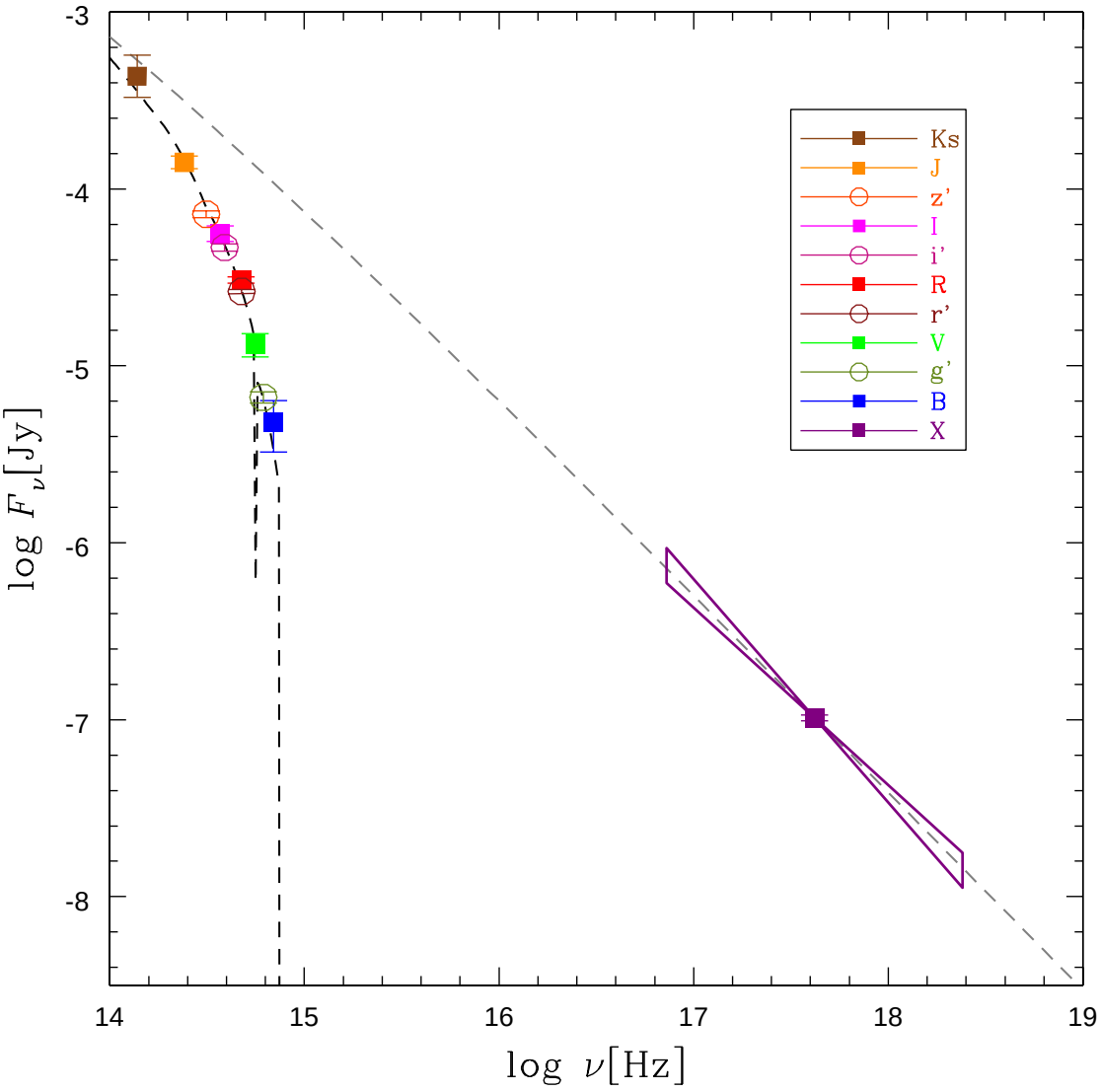

Figure 4.2 Fitted constant-density ($k = 0$) model NIR–X-ray spectrum (dashed black curve) for GRB 090313, at the jet break time $t = t_{23} \approx 1$ day. The dashed gray curve shows the best-fit intrinsic emission spectrum. All data points at times $\log t > -0.7$ in each filter were scaled to $t = t_{23}$ using the best-fit model and averaged (for plotting purposes only). The triangular regions about the averaged X-ray data point illustrate the 0.3–10 keV band range, and the 1σ uncertainty in the mean effective spectral index $\overline{\beta}_X^{\rm eff} = -1.14^{+0.14}_{-0.12}$.

and the parameters describing dust extinction in the Milky Way were linked across both time slices. Also linked were: the times of the light curve peak and jet break, $\log t_{12}$ and $\log t_{23}$; the smoothing parameters of the peak and jet break, s_{12} and s_{23}; the electron energy index p; the circumburst density index k; the optical/NIR and X-ray slops, $\sigma^O_{\log F_\nu}$ and $\sigma^X_{\log F}$; and all group calibration offsets $\Delta_{c,n}$.

The parameters that were allowed to freely vary between the two time slices, singly or in groups, were: the flux normalization, $\log F_0$, and the frequency of the cooling break at the time of the light curve peak, ν_{c0}; the source-frame dust extinction parameters, A_V, c_2, c_4, and the secondary parameters (constrained by priors) that with c_2 determine the source-frame extinction parameters c_1, R_V, BH, x_0 and γ; and the neutral hydrogen column density N_H of the source-frame damped Lyα absorber. To obtain a "baseline" model, to which the relative fitnesses of various subparameterizations of the model can be compared, we first obtained a fit with all of the above parameters unlinked across the two time slices, and with the circumburst density index k allowed to vary freely. This fit settled on a value of $k = -0.35$, intermediate between the standard uniform-density ISM ($k = 0$) and stellar wind ($k = -2$) models, with $-2\ln\mathcal{P}$ equal to a best-fit value $(-2\ln\mathcal{P})_{\rm min}$, where $\mathcal{P}$ is the posterior probability of the model, given the data and all prior constraints (§3.5.2). Note that the value of $(-2\ln\mathcal{P})_{\rm min}$ does not provide an absolute measure of fitness. However, for a given data set, model parameterization, and set of priors, the relative likelihood of various subparameterizations of the model can be quantified by the difference in their fitnesses, $\Delta(-2\ln\mathcal{P}) = (-2\ln\mathcal{P}) - (-2\ln\mathcal{P})_{\rm min}$; the relative likelihood of one subparameterization with respect to the global minimum is given by $e^{-\frac{1}{2}\Delta(-2\ln\mathcal{P})}$ (§2.6.2).

We then ran fits with k fixed at 0 and -2 (i.e., ISM and wind models), and found these fits were not significantly worse than the $k = -0.35$ global minimum, with $\Delta(-2\ln\mathcal{P}) = 0.10$ and 2.92, respectively. Since the vast majority of afterglow fits

in the literature assume either $k = 0$ or $k = -2$, we decided to further explore only these two cases. We note, however, that the inability to obtain statistically significant constraints on the value of the circumburst density index k is primarily due to the fact that the position of the cooling break, and in particular its direction of motion, are not well constrained by this data set. In principle, with sufficient time and frequency coverage, it should be possible to constrain arbitrary radial circumburst medium density profiles from GRB afterglow observations using our modeling techniques, and even detect deviations from a single $n(r) \propto r^k$ profile, e.g., due to variable stellar winds from the collapsar progenitor.

Table 4.1 summarizes $\Delta(-2 \ln \mathcal{P})$ for ISM and wind models for various linkings of parameters across the two timeslices. It is immediately obvious that if we link all emission, extinction and absorption parameters across the two time slices (i.e., if we fit to all the data with a single set of parameters), the change in fitness, $\Delta(-2 \ln \mathcal{P}) = 22.89$ and 36.23 for the $k = 0$ and $k = -2$ models, respectively, is unacceptably large. We explore two basic possibilities: either the rebrightening is due to an *intrinsic* change in the emission of the GRB afterglow, or it is due to a change in the *environmental* properties of the circumburst medium through which the radiation emitted by the afterglow travels. In the former case, the parameters that we allow to vary are the flux normalization $\log F_0$ and the frequency of the cooling break $\log \nu_{c0}$ at the fitted light curve peak time t_{12} and (fixed) reference frequency ν_0. With these quantities, and the spectral and temporal indices predicted by the values of p and k (which we assume do not vary between the two time slices), the entire emission spectrum can be constructed. In the latter case, we can vary the intrinsic extinction parameters A_V, c_2, c_4, c_1, R_V, BH, x_0 and γ, and/or the neutral hydrogen column density N_H of the damped Lyα absorber. Note that some of the extinction parameters are actually functions of secondary parameters that are constrained by priors, and, in the case of

Table 4.1. Relative Fitness $\Delta(-2\ln\mathcal{P})$ for GRB 090313 Model Fits

	Varying Parameters[a]					
	Em, Ext, Abs	None	Em	Ext, Abs	Ext	Abs
Linked	None	Em, Ext, Abs	Ext, Abs	Em	Em, Abs	Em, Ext
$k = 0$	0.10	22.89	1.41	2.48	2.56	22.47
$k = -2$	2.92	36.23	5.04	10.91	10.94	35.99

[a]Em = $\log F_0$, $\log \nu_{c0}$; Ext = A_V, c_2, c_4, c_1, R_V, BH, x_0, γ; Abs = N_H.

c_1, R_V and BH, are also constrained by c_2 (§3.3.4).

From Table 4.1, it is obvious that the observed rebrightening cannot be attributed solely to a varying of N_H ("Abs"). This is not surprising, given that the rebrightening is observed in all optical/NIR bands, while the damped Lyα absorption line affects only V and g' observations. The most probable explanation is variable intrinsic emission ("Em"). However, the fits with varying extinction and absorption ("Ext, Abs") or extinction alone ("Ext") are not significantly worse, based solely on their fitnesses, than the varying emission fits, particularly in the constant-density circumburst medium ($k = 0$) case. However, the only way to achieve a simultaneous brightening in both the NIR and optical fluxes by varying extinction alone is for A_V to decrease at late times (in the case of the $k = -2$ model, to decrease rather dramatically to nearly zero) and for c_2 to increase dramatically (from $c_2 \approx 2$ to $c_2 \approx 6$ for $k = 0$ and to $c_2 \approx 40$ for $k = -2$). We consider this scenario to be physically implausible. Any change in the observed extinction is expected to occur as the jet widens with time, illuminating populations of dust that are progressively further off-axis from the initial line of sight to the GRB. At early times, the dust along the line of sight is expected to have been modified or destroyed by the GRB and afterglow emission itself. Fragmentation of larger dust

grains would result in proportionally greater source-frame UV extinction, and hence higher measured values of c_2 along the line of sight, while complete destruction of dust grains would result in lower values of A_V. At later times, the expanding jet is expected to increasingly illuminate unmodified populations of dust. Thus, if we see any change in emission over time, it would most likely be from *lower* A_V and *higher* c_2 to *higher* A_V and *lower* c_2, not the other way around.

We therefore conclude that *the most likely explanation for the brightening is a change in the intrinsic emission, not a change in extinction.* It is this case that we explore further here. We computed error bar marches to obtain fitted projected probability distributions (§2.6.4) for the fully free model parameters, i.e., for parameters that are not functionally linked to other parameters or constrained by priors; the uncertainties quoted in Table 4.2 were computed from these fitted distributions. Furthermore, as we discuss below in §4.5, from the fitted changes in $\log F_0$ and $\log \nu_{0c}$ it is possible to compute potential changes in macro- or micro-physical properties ($E_{\rm iso}$, ϵ_e, ϵ_B and n or A_*) of the afterglow and its environment that could explain the changes in the emitted spectrum and, in principle, to rule out one or more of these explanations.

The best-fit free emission model parameters ($\log F_0$, $\log \nu_{c0}$, $\log t_{12}$, s_{12}, $\log t_{23}$, s_{23}, p, $\Delta\alpha_1$, $\sigma^O_{\log F_\nu}$ and $\sigma^X_{\log F}$) and free source-frame extinction/absorption parameters (A_V, c_2, c_4 and $\log N_H$), and their fitted projected uncertainties are presented in Table 4.2 for this intrinsic brightening model, for both $k = 0$ and $k = -2$ cases. The best-fit functionally linked emission parameters (the spectral and temporal indices α and β in the various regimes) are presented in Table 4.3, with uncertainties propagated from the fitted uncertainties in the electron energy index p. Table 4.4 presents the best-fit prior-constrained extinction parameters ($R_V^{\rm MW}$, x_0 and γ), and the best-fit correlative prior-constrained extinction parameters (c_1, R_V and BH). The former are functions of secondary parameters that describe the best-fit model probability distributions of

each parameter (§§3.3.2 & 3.3.4.4). The latter are functions of the free parameter c_2, and of secondary parameters that describe the best-fit model distributions fit to their correlations with c_2 (§§3.3.4.1, 3.3.4.2 & 3.3.4.3); the quoted uncertainties are propagated from the fitted uncertainties in c_2, using the best-fit secondary correlative parameter values. Figures 4.3 & 4.4 show the best-fit spectral models, referenced to $\log t = -0.7$ (the time at which we divide the data into two time slices) for the $k = 0$ and $k = -2$ fits. Figures 4.5 & 4.6 show the best-fit light curves, plotted against the observed data, for the $k = 0$ and $k = -2$ models.

The best-fit source frame extinction curves have $A_V \simeq 0.4$, $c_2 \simeq 2.2$ and $c_4 \simeq 0.3$ for $k = 0$, and $A_V \simeq 0.3$, $c_2 \simeq 2.3$ and $c_4 \simeq 0.4$ for $k = -2$. These values are consistent with either: 1) SMC-like dust, typical of older star-forming regions where supernova shocks have fragmented dust grains; or 2) GRB/afterglow-modified dust, where the prompt emission from the burst fragments the dust along the line of sight. In either scenario, the fragmentation of larger grains into smaller grains is expected to increase the measured values of c_2 and, possibly, c_4, compared to dust typical in the diffuse ISM.

Because the damped Lyα absorption line is a narrow spectral feature, it is difficult to accurately measure the column density of neutral hydrogen, N_H, based solely on photometric observations. However, we are able to place 3σ upper limits $N_H < 6.3 \times 10^{21}$ cm^{-2} for $k = 0$ and $N_H < 4.0 \times 10^{21}$ cm^{-2} for $k = -2$. Note, however, that the effective source-frame column density obtained by the *Swift* XRT model is $\overline{N}_{H_X}^{\rm eff} = 2.99^{+0.77}_{-0.71} \times 10^{22}$ cm^{-2}, a factor of $\approx$ 5–8 times greater than our 3σ upper limits. There are two possible explanations for this discrepancy. The first explanation is that since X-ray absorption at these frequencies is primarily due to metals, and a solar metal-to-gas ratio is assumed in the model, the large value of $\overline{N}_{H_X}^{\rm eff}$ could be due to the source-frame absorbing gas having a higher than solar metallicity. However, since GRB 090313 occurred at a redshift $z_{\rm GRB} = 3.375$, we consider this scenario unlikely.

The second explanation is that since our inferred value of N_H is based on modeling an absorption line due to neutral hydrogen, the lower values that we obtain could be due to the source-frame hydrogen being ionized at the $\approx$ 80–90% level along the line of sight (or greater, in sub-solar metallicity is assumed). We consider this the more likely explanation. If correct, it can best be explained by ionization by the prompt GRB/afterglow emission itself, which would imply that most of the gas is local to the GRB. Indeed, $\overline{N}_{H_X}^{\rm eff} = 3 \times 10^{22}$ cm^{-2} is consistent with what would be expected if GRB 090313 occurred in a giant molecular cloud (GMC; e.g., Reichart & Price, 2002). Given the absence of a bright optical flash for this burst, which would be required to sublimate or completely destroy the dust on GMC scales, this implies that the dust is likely local as well, in which case the large values of c_2 might best be explained by the fragmentation of dust along the line of sight (e.g., Draine, 2000; Fruchter et al., 2001; Reichart & Price, 2002).

Table 4.2. Best-Fit Free Model Parameters for GRB 090313, with $\log F_0$ and $\log \nu_{c0}$ Allowed to Vary Between Timeslices

	ISM model ($k = 0$)		Wind model ($k = -2$)	
	Slice 1	Slice 2	Slice 1	Slice 2
Emission Parameters[a], $\log \nu_0 \equiv 14.5948$				
$\log F_0$	-1.67 ± 0.16	$-1.24^{+0.16}_{-0.26}$	$-1.92^{+0.22}_{-0.12}$	$-1.55^{+0.21}_{-0.15}$
$\log \nu_{C,0}$	$15.04(> 12.64)$	$14.56(< 15.87)$	$11.18(< 13.58)$	$11.86(< 13.50)$
$\log t_{12}$	$-2.003^{+0.042}_{-0.065}$		$-1.995^{+0.047}_{-0.084}$	
s_{12}	$1.03^{+0.40}_{-0.31}$		$0.90^{+0.39}_{-0.31}$	
$\log t_{23}$	$0.0125^{+0.064}_{-0.051}$		$-0.036^{+0.045}_{-0.040}$	
s_{23}	(> 4.0)		(> 3.7)	
p	$2.236^{+0.105}_{-0.091}$		$2.129^{+0.066}_{-0.058}$	
$\Delta\alpha_1$	$2.77^{+0.39}_{-0.26}$		$2.89^{+0.55}_{-0.27}$	
$\sigma^O_{\log F_\nu}$	0.067		0.068	
$\sigma^X_{\log F}$	0.000		0.000	
Source-Frame Extinction/Absorption Parameters[a], $z_{\rm GRB} \equiv 3.375$				
A_V	$0.386^{+0.089}_{-0.124}$		$0.327^{+0.087}_{-0.135}$	
c_2	$2.20^{+1.11}_{-0.38}$		$2.27^{+1.63}_{-0.43}$	
c_4	$0.32^{+0.13}_{-0.11}$		$0.42^{+0.25}_{-0.17}$	
$\log N_H$	$20.8(< 21.8)$		$20.8(< 21.6)$	

[a] $\pm 1\sigma$ uncertainties; 3σ upper or lower limits in parentheses.

Table 4.3. Best-Fit Functionally Linked Emission Parameters for GRB 090313, with $\log F_0$ and $\log \nu_{c0}$ Allowed to Vary Between Timeslices

	ISM model ($k = 0$)	Wind model ($k = -2$)
Emission Parameters,[a] $\log \nu_0 \equiv 14.5948$		
α_{a1}	$1.84^{+0.39}_{-0.26}$	$1.54^{+0.55}_{-0.27}$
α_{b1}	$1.59^{+0.39}_{-0.26}$	$1.67^{-0.55}_{-0.27}$
α_{a2}	$-0.927^{-0.079}_{+0.068}$	$-1.346^{-0.049}_{+0.043}$
α_{b2}	$-1.177^{-0.079}_{+0.068}$	$-1.096^{-0.049}_{+0.043}$
α_{a3}	$-2.236^{-0.105}_{+0.091}$	$-2.129^{-0.066}_{+0.058}$
α_{b3}	$-2.236^{-0.105}_{+0.091}$	$-2.129^{-0.066}_{+0.058}$
$\beta_{a1} = \beta_{a2} = \beta_{a3}$	$-0.618^{-0.053}_{+0.045}$	$-0.564^{-0.033}_{+0.029}$
$\beta_{b1} = \beta_{b2} = \beta_{b3}$	$-1.118^{-0.053}_{+0.045}$	$-1.064^{-0.028}_{+0.027}$

[a] $\pm 1\sigma$ uncertainties obtained by propagation of fitted uncertainty in electron energy index p.

Table 4.4. Best-Fit Prior-Constrained and Functionally Linked Extinction Parameters for GRB 090313, with $\log F_0$ and $\log \nu_{c0}$ Allowed to Vary Between Timeslices

	ISM model ($k = 0$)	Wind model ($k = -2$)
Extinction/Absorption Parameters[a], $z_{\rm GRB} \equiv 3.375$, $E(B-V)_{\rm MW} \equiv 0.02793$		
$R_{V,\rm MW}$	2.642	2.642
c_1	$-4.7^{-3.6}_{+1.2}$	$-4.9^{-5.4}_{+1.4}$
R_V	$2.797^{-0.069}_{+0.013}$	$2.822^{-0.107}_{+0.031}$
BH	$0.53(< 1.74)$	$0.38(< 1.86)$
x_0	4.584	4.584
γ	0.894	0.894

[a]$\pm 1\sigma$ uncertainties in source-frame R_V & c_1 and 3σ upper limits on BH obtained by propagation of fitted uncertainty in c_2, using best-fit secondary correlation model parameters.

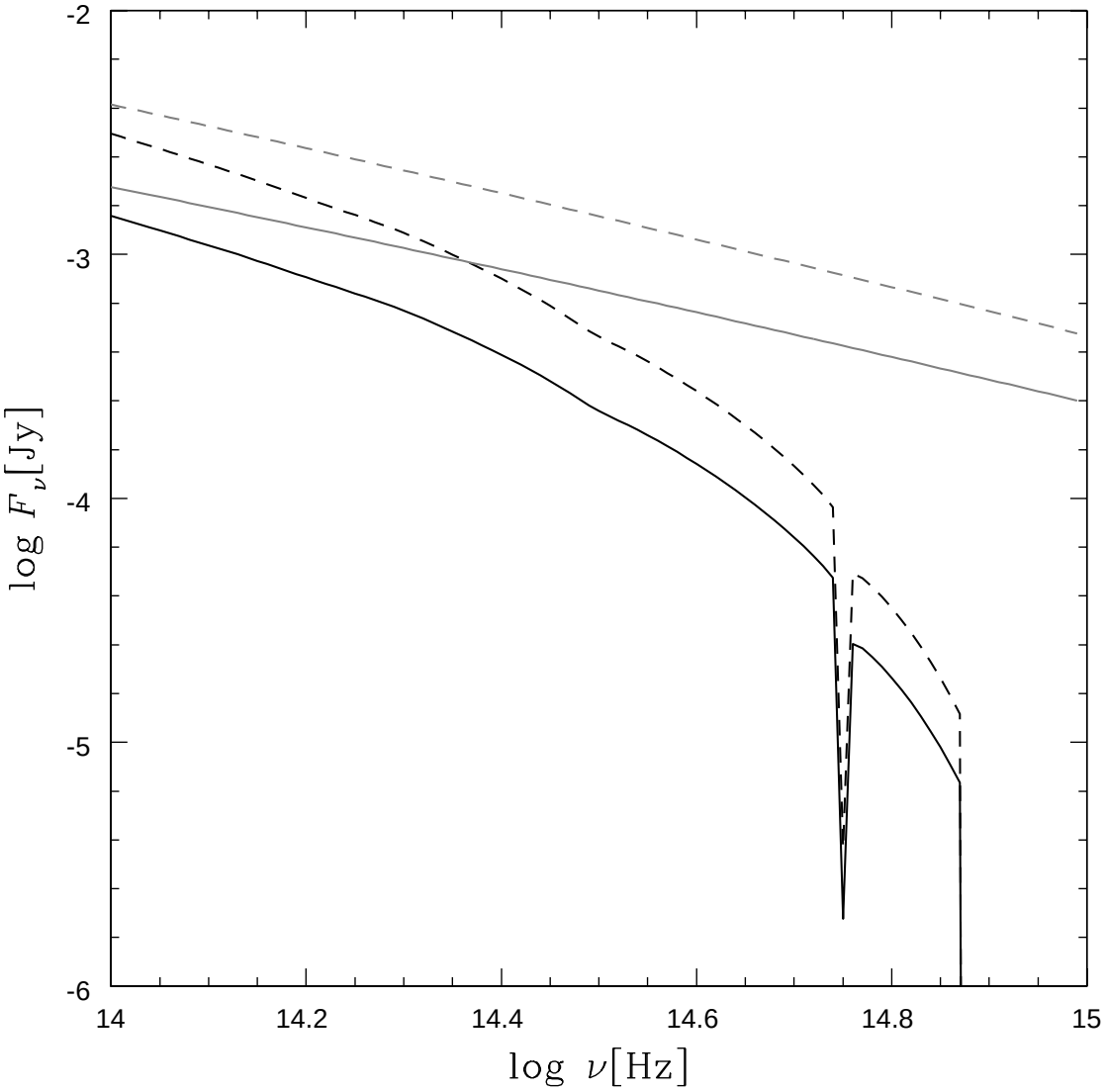

Figure 4.3 Fitted constant-density ($k = 0$) model spectra for GRB 090313, at $\log t = -0.7$. Data were grouped into two time slices, divided at $\log t = -0.7$; all parameters describing extinction and absorption were linked across the two slices, while the flux normalization and cooling break frequencies were allowed to vary. Solid and dashed black curves show the best-fit spectra, with extinction and absorption included, for time slices 1 and 2, respectively. Solid and dashed gray curves show the best-fit intrinsic emission spectra.

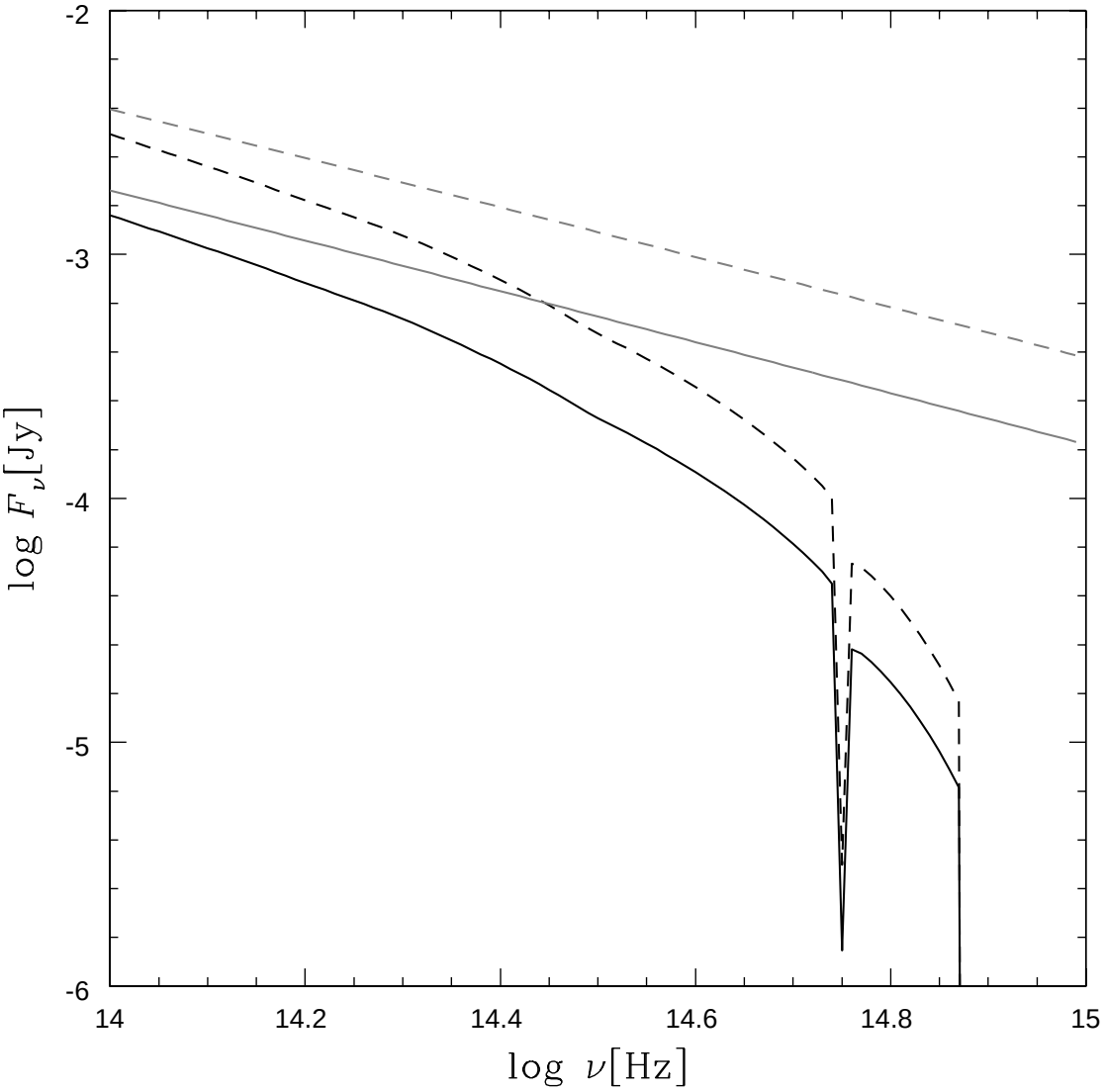

Figure 4.4 Fitted stellar wind ($k = -2$) model spectra for GRB 090313, at $\log t = -0.7$. Data were grouped into two time slices, divided at $\log t = -0.7$; all parameters describing extinction and absorption were linked across the two slices, while the flux normalization and cooling break frequencies were allowed to vary. Solid and dashed black curves show the best-fit spectra, with extinction and absorption included, for time slices 1 and 2, respectively. Solid and dashed gray curves show the best-fit intrinsic emission spectra.

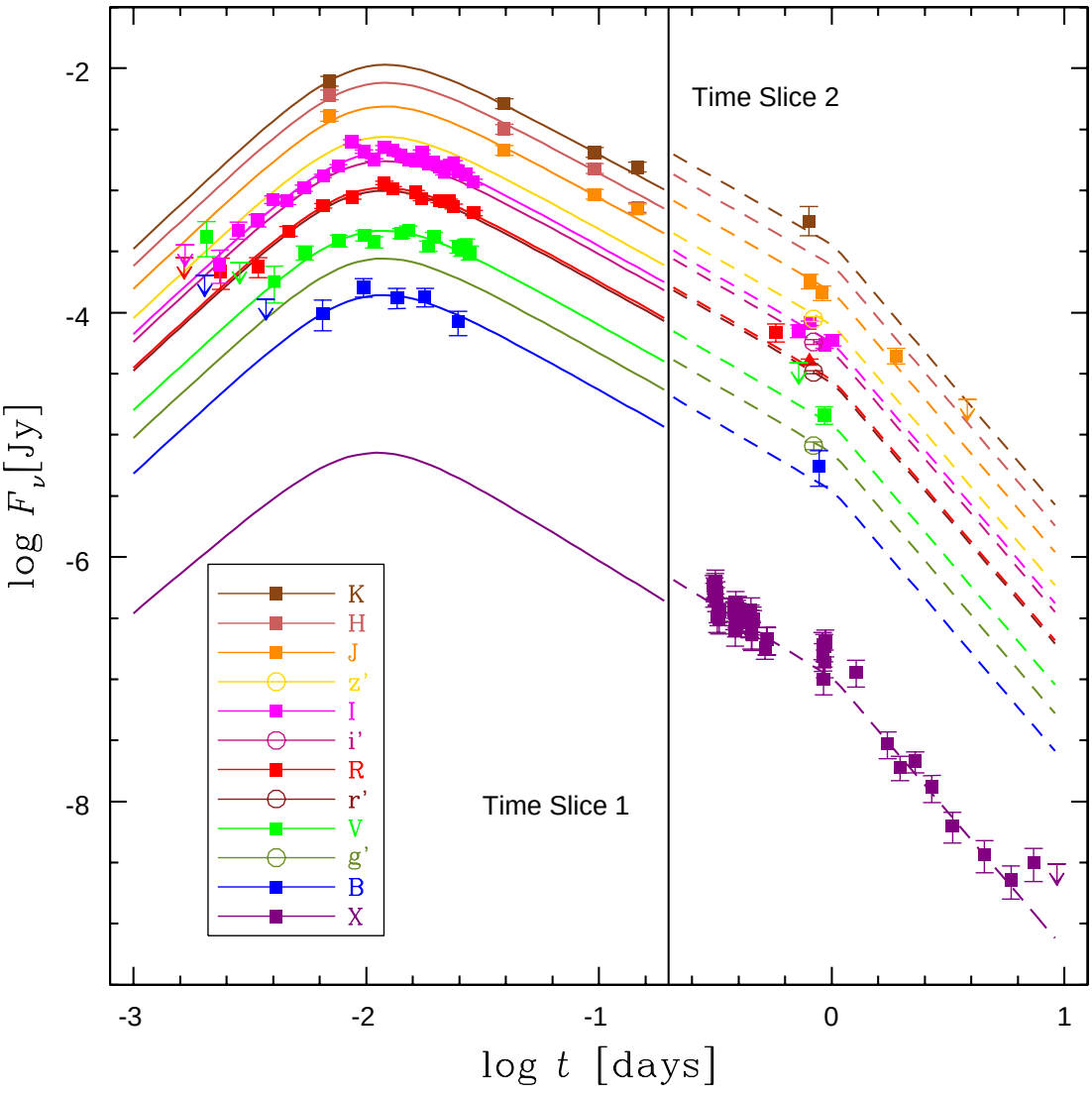

Figure 4.5 Observed flux densities and fitted constant-density ($k = 0$) model NIR–X-ray light curves for GRB 090313. Data were grouped into two time slices, divided at $\log t = -0.7$; all parameters describing extinction and absorption were linked across the two slices, while the flux normalization and cooling break frequencies were allowed to vary. Solid and dashed lines show the best-fit light curves for time slices 1 and 2, respectively. NIR and optical model light curves are weighed averages over the response function of each photometric filter. The X-ray light curve is plotted for the mean log frequency of the $0.3 - 10$ keV band. The integrated X-ray flux data have been scaled to the effective model flux density at this frequency (Equation 4.18).

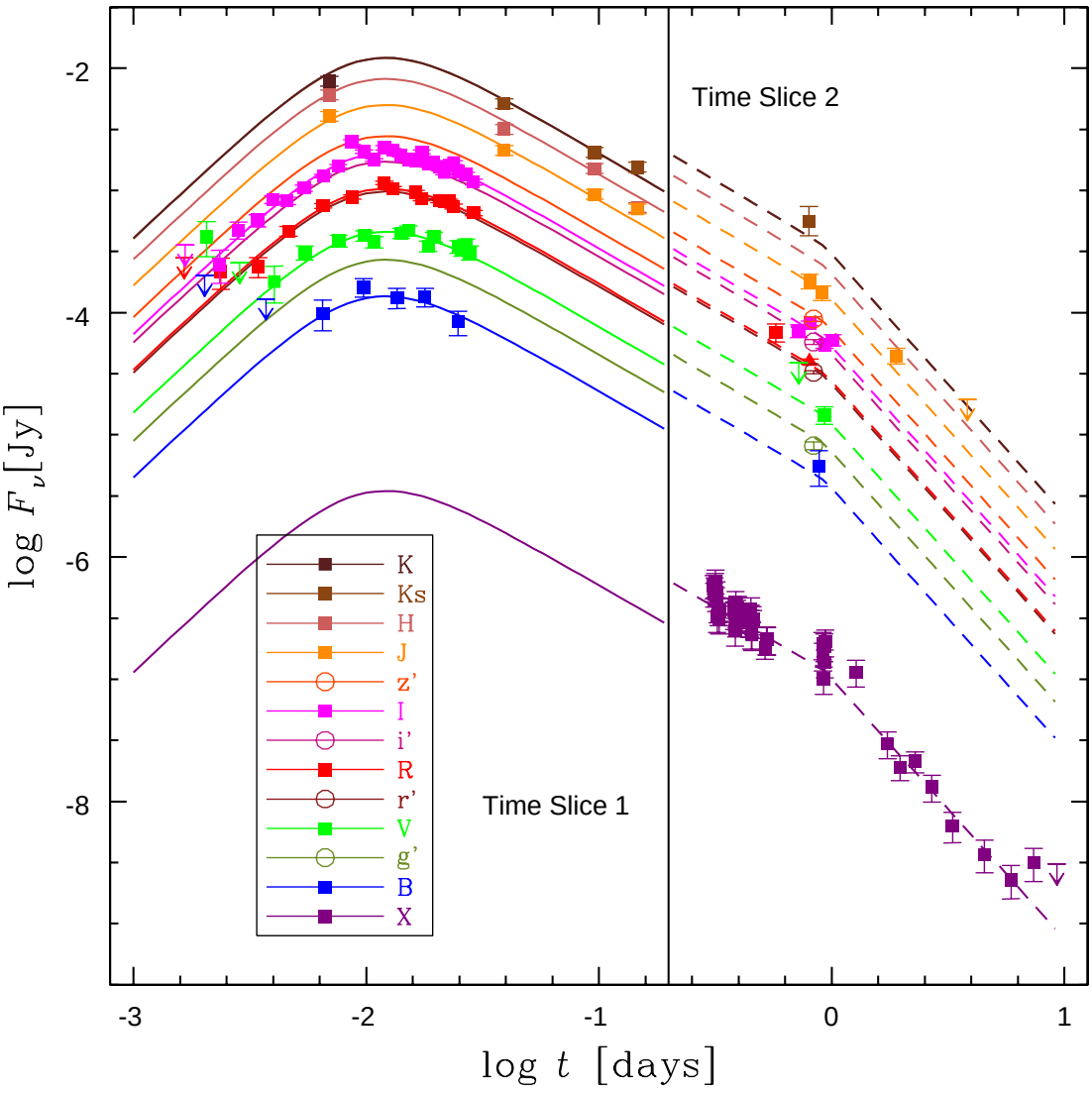

Figure 4.6 Observed flux densities and fitted stellar wind ($k = -2$) model NIR–X-ray light curves for GRB 090313. Data were grouped into two time slices, divided at $\log t = -0.7$; all parameters describing extinction and absorption were linked across the two slices, while the flux normalization and cooling break frequencies were allowed to vary. Solid and dashed lines show the best-fit light curves for time slices 1 and 2, respectively. NIR and optical model light curves are weighed averages over the response function of each photometric filter. The X-ray light curve is plotted for the mean log frequency of the $0.3 - 10$ keV band. The integrated X-ray flux data have been scaled to the effective model flux density at this frequency (Equation 4.18).

4.5 Energy Injection or Density Variation: *Galapagos*-Enabled Science

The flexibility of our genetic algorithm software package *Galapagos* to allow simultaneous fitting to data groups (e.g., time slices) with parameter linking enables us to easily test hypotheses that would otherwise be extremely difficult or impossible to model. As described above, for GRB 090313, we are able to quantify the relative probability that the observed brightening of the afterglow is due to variable extinction and absorption, or due to variable intrinsic emission. While variable extinction is not ruled out statistically, it requires fitted extinction parameter values that we deem to be physically implausible. The simplest, and most plausible, explanation is that there was an intrinsic rebrightening of the afterglow emission at some time $\log t \approx -0.7$ days, which we model with *Galapagos* by unlinking the parameters describing the flux density normalization, $\log F_0$, and the frequency of the cooling break, $\log \nu_{c0}$, while holding all other parameters linked across the two time slices.

In the standard blast wave model for synchrotron emission from relativistic shocks in an ultrarelativistic jet (e.g., Mészáros & Rees, 1997; Sari et al., 1998; Chevalier & Li, 1999), fitted changes in the flux normalization of an afterglow spectrum and in the frequency of the cooling break can be explained by proportional changes in a number of physical quantities for which theoretical scaling relations have been derived: the isotropic-equivalent energy of the blast wave, $E_{\rm iso}$; the fraction of this energy that is distributed into accelerated electrons, ϵ_e, and into the magnetic field, ϵ_B; and the circumburst density normalization, n (for a constant-density $k = 0$ model) or A_* (for a wind-driven $k = -2$, or other non-constant density model). Sari et al. (1998) present the theoretical scaling relations for the $k = 0$ case, and Chevalier & Li (1999) present them for the $k = -2$ case.

In either case, the (unsmoothed) observed flux density at a frequency ν at any given time is related to the flux density at the synchrotron peak F_{ν_m}, the frequency of the synchrotron peak, ν_m, the frequency of the cooling break ν_c, and the electron energy index p:

$$\begin{aligned} F_\nu &= F_{\nu_m} \left(\frac{\nu}{\nu_c}\right)^{(p+1)/2} \text{ if } \nu < \nu_c \,, \\ F_\nu &= F_{\nu_m} \left(\frac{\nu_c}{\nu_m}\right)^{(p+1)/2} \left(\frac{\nu}{\nu_c}\right)^{p/2} \text{ if } \nu > \nu_c \,. \end{aligned} \tag{4.20}$$

Assuming adiabatic evolution of the blast wave (a condition that is met as soon as ν_m drops below ν_c, which typically occurs within seconds), for a constant density circumburst medium ($k = 0$) of density n:

$$\begin{aligned} \nu_c &\propto \epsilon_B^{-3/2} E_{\rm iso}^{-1/2} n^{-1} t^{-1/2} \,, \\ \nu_m &\propto \epsilon_e^2 \epsilon_B^{1/2} E_{\rm iso}^{1/2} t^{-3/2} \,, \\ F_{\nu_m} &\propto \epsilon_B^{1/2} E_{\rm iso} n^{1/2} \,. \end{aligned} \tag{4.21}$$

For a wind-driven ($k = -2$) medium with density normalization A_*:

$$\begin{aligned} \nu_c &\propto \epsilon_B^{-3/2} E_{\rm iso}^{1/2} A_*^{-2} t^{1/2} \,, \\ \nu_m &\propto \epsilon_e^2 \epsilon_B^{1/2} E_{\rm iso}^{-2} t^{1/2} \,, \\ F_{\nu_m} &\propto \epsilon_B^{1/2} E_{\rm iso}^{1/2} A_* t^{-1/2} \,. \end{aligned} \tag{4.22}$$

Substituting the scaling relations of Equations 4.21 into Equations 4.20 gives the expected scaling of F_ν for $k = 0$:

$$F_\nu \propto \epsilon_e^{-p-1} \epsilon_B^{(-p+1)/4} E_{\rm iso}^{(-p+3)/4} n^{1/2} t^{(3p+1)/4} \nu^{(p+1)/2} \text{ if } \nu < \nu_c \,,$$

$$F_\nu \propto \epsilon_e^{(-p-1)} \epsilon_B^{(-p-2)/4} E_{\rm iso}^{(-p+2)/4} t^{(3p-1)/4} \nu^{p/2} \text{ if } \nu > \nu_c \,. \quad (4.23)$$

Substituting the scaling relations of Equations 4.22 into Equations 4.20 gives the expected scaling of F_ν for $k = 2$:

$$\begin{aligned} F_\nu &\propto \epsilon_e^{-p-1} \epsilon_B^{(-p+1)/4} E_{\rm iso}^{(-p+1)/4} A_* t^{(3p+1)/4} \nu^{(p+1)/2} \text{ if } \nu < \nu_c \,, \\ F_\nu &\propto \epsilon_e^{(-p-1)} \epsilon_B^{(-p-2)/4} E_{\rm iso}^{(-p+2)/4} t^{(3p+2)/4} \nu^{p/2} \text{ if } \nu > \nu_c \,. \end{aligned} \quad (4.24)$$

With these scaling relations in Equations 4.23 & 4.24, we can compute what physical change in any of these quantities ($\Delta \log E_{\rm iso}$, $\Delta \log \epsilon_e$, $\Delta \log \epsilon_B$ or $\log n$ or $\log A_*$) would be necessary to explain, alone, our fitted change in the flux density normalization $\Delta \log F_0$. Similarly, using the scaling relations for ν_c in Equations 4.21 & 4.22, we can compute what physical change in any of these quantities (except $\log \epsilon_e$) would be necessary to explain, alone, the fitted change in the frequency of the cooling break frequency $\log \nu_{c0}$. However, in the fits presented in Tables 4.2 & 4.3, while we do arrive at a best-fit value for ν_{c0}, error bar marches reveal that the data are actually only sufficient to provide upper or lower limits on $\log \nu_{c0}$. Hence, we confine our analysis to the fitted changes in $\log F_0$. But it is important to note that, if both $\Delta \log F_0$ and $\Delta \log \nu_{c0}$ were measured to a sufficient degree of statistical certainty, it would be possible, in principle, to compare the predictions of their respective scaling relations, and to rule out one of more of the physical changes on the basis of consistency.

Since the best-fit values of and fitted projected uncertainties in the free parameter $\log F_0$ are highly correlated between the two time slices, we fit instead to the parameter $\Delta \log F_0$, i.e., the difference in flux normalization between the two time slices. The best-fit values for the $k = 0$ and $k = -2$ models are $\Delta \log F_0 = 0.43 \pm 0.07$ and $\Delta \log F_0 = 0.37 \pm 0.06$, respectively. In these, and in the previous fits, $\log \nu_0 \lesssim \log \nu_{c0}$ for the $k = 0$ model, and $\log \nu_0 > \log \nu_{c0}$ for the $k = -2$ model. We therefore compute

the possible changes in the various physical parameters using the scaling relations for $\nu < \nu_c$ in the former case, and for $\nu > \nu_c$ in the latter case (Equations 4.23 & 4.24). We present the changes in the physical quantities and their 1σ uncertainties in Table 4.5. Note that, since the reference frequency $\nu_0 > \nu_{c0}$ for the $k = -2$ fit, we can place no constraints on A_* using $\Delta \log F_0$ (Equation 4.24). For the $k = 0$ case, the fraction of energy in the electrons, ϵ_e, or in the magnetic field, ϵ_B, would have to increase by factors of $\approx$2 or $\approx$3, respectively, to explain the observed $\Delta \log F_0$. For the $k = -2$ case, ϵ_e and ϵ_B would have to increase by factors of $\approx$2.5 and $\sim 10^{11}$. However, we consider sudden (and, especially, dramatic) changes in the partition of energy in electrons and in the magnetic field to be physically implausible (also, depending on the actual pre-change values of ϵ_e and ϵ_B, changes by factors of a few may result in either being $>$1, which is physically meaningless.) The two more plausible sources of intrinsic brightening are *energy injection* ($\Delta \log E_{\rm iso}$, e.g., due to collisions of later-time ejecta with the decelerating shock), or *density variation* ($\Delta \log n$ or $\Delta \log A_*$, e.g., due to variable stellar wind profiles from the GRB progenitor). For both the $k = 0$ and $k = -2$ cases, a increase in $E_{\rm iso}$ by a factor of $\approx$2.1 and $\approx$2.5, respectively, would be sufficient to explain the observed brightening. For the $k = 0$ case, a discontinuous transition from one constant-density medium to another that is $\approx$7 times denser would also be sufficient to explain the brightening, while, for the $k = -2$ case, the standard blast wave model predicts no rebrightening at $\nu_0 > \nu_{c0}$ if A_* changes (Equation 4.24).

While we can state no firm conclusions on possibile changes in physical parameters as the cause of the observed rebrightening in the afterglow of GRB 090313, it should be clear that, in principle, by carefully modeling observations of GRB afterglows, using the data grouping and parameter linking features of *Galapagos*, it is possible to confirm (or rule out) one or both of these scenarios. If, for instance, we were better able to constrain the change in the cooling break frequency, $\Delta \log \nu_{c0}$ (and to determine

Table 4.5. Changes in Individual Physical Quantities Inferred from Fitted $\Delta \log F_0$ for GRB 090313

Quantity	$k = 0$	$k = -2$
$\Delta \log F_0$	0.43 ± 0.07	0.37 ± 0.06
Micro-physical Parameters		
$\Delta \log \epsilon_e$	0.35 ± 0.06	0.41 ± 0.06
$\Delta \log \epsilon_B$	0.53 ± 0.09	11.41 ± 1.98
Macro-physical Parameters		
$\Delta \log E_{\rm iso}$	0.33 ± 0.06	0.39 ± 0.06
$\Delta \log n$	0.86 ± 0.15	$\cdots$ [a]

[a]For $k = -2$ and $\nu_0 > \nu_{c0}$, F_0 does not scale with A_* (Equation 4.24).

whether $k = 0$, $k = -2$ or some other density profile was favored), we could use its scaling relations (Equations 4.21 & 4.22) as a consistency check on the predictions of the $\Delta \log F_0$ scaling relations. Finally, we note that if a data set is of sufficient time and frequency coverage to permit determination of the the flux at the synchrotron peak, F_{ν_m}, and the frequencies of the synchrotron peak, cooling break and self-absorption break, ν_m, ν_c and ν_a, respectively, it is possible, in principle, to calculate values, instead of possible changes in values, for each of these physical quantities. Such calculations will be an active area of future research as the Afterglow Modeling Project progresses.

Appendix A

Convolution of Two Asymmetric Gaussian Distributions

A.1 Asymmetric 1D Probability Distributions

In §2.2.1 we discuss the general problem of computing the joint probability of a model distribution, $p^{\rm mod}$, and a single measurement with an intrinsic probability distribution $p_n^{\rm int}$. Here, we consider the simpler case of a 1D statistic (we discuss the treatment of asymmetric distributions in a 2D statistic in Appendix B). The model distribution is described by some curve $y_c(x;\vartheta_m)$, which is a function of the independent variable x and some set of parameters $\{\vartheta_m\}$, and by some probability density function $p^{\rm mod}(y;y_c(x;\vartheta_m))$ about that curve. The measured data point is described by some value, y_n, and a probability density function $p_n^{\rm int}(y;y_n)$. The joint probability of the model distribution and the measurement is given by:

$$p_n = \int_{-\infty}^{\infty} p^{\rm mod}(y;y_c(x;\vartheta_m))p_n^{\rm int}(y;y_n)dy\,. \tag{A.1}$$

This integral may or may not yield an analytic expression for p_n, depending on the functional forms of the two probability density functions.

The most familiar, and simplest, scenario is when both probabilities are normally distributed, i.e., $p^{\rm mod}(y;y_c(x;\vartheta_m)) = G(y;y_c(x;\theta),\sigma_y)$ and $p_n^{\rm int}(y;y_n) = G(y;y_n,\sigma_{y,n})$,

where G is the Gaussian function:

$$G(y; y_n, \sigma_{y,n}) \equiv \frac{1}{\sqrt{2\pi}\sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{y - y_n}{\sigma_{y,n}}\right)^2\right]. \tag{A.2}$$

In this case, the joint probability integral of Equation A.1 also has a Gaussian form:

$$p_n = \frac{1}{\sqrt{2\pi}\Sigma_{y,n}} \exp\left[-\frac{1}{2}\left(\frac{y_n - y_c(x; \vartheta_m)}{\Sigma_{y,n}}\right)^2\right], \tag{A.3}$$

where $\Sigma_{y,n}$ is the sum in quadrature of the sample variance, or "slop", and the intrinsic uncertainty:

$$\Sigma_{y,n} \equiv \sqrt{\sigma_y^2 + \sigma_{y,n}^2}\,. \tag{A.4}$$

The joint probability integral can equivalently be expressed in terms of a convolution of a zero-mean Gaussian $G(y; 0; \sigma_y)$ and the data point's intrinsic Gaussian pdf, evaluated at $y = y_c(x; \vartheta_m)$:

$$p_n = \int_{-\infty}^{\infty} \left[G(y; 0, \sigma_y) * G(y; y_n, \sigma_{y,n})\right] \delta(y - y_c(x; \vartheta_m)) dy\,, \tag{A.5}$$

where $\delta(y - y_c(x; \vartheta_m))$ is the Kronecker delta function.

There are many instances where one or both probability distribution functions are not normally distributed. For instance, when fitting to GRB afterglows, we evaluate the statistic in linear flux space, but assume the model distribution is log-normally distributed. Similarly, while proper flux measurement uncertainties from a CCD image should be roughly normally (or Poisson) distributed in linear flux space (a CCD camera is an intrinsically linear flux-measuring device), we often must include in our data sets measurements that have less accurately been quoted with symmetric error bars in log flux (or magnitude) space. There are also situations where published data are quoted with asymmetric 1σ error bars, but with no obvious information as to the form of the

underlying probability distribution.

Regardless, these data must be included in our analysis somehow. In the absence of other information, we must transform these log-normal distributions into linear space as best we can. One common practice is to compute a "symmetrized" distribution. If the quoted measurement and its uncertainty are $\log F_n(+\sigma_{\log F_n +}, -\sigma_{\log F_n -})$, then the transformed value may be approximated as:

$$F_n = 0.5 \left(10^{\log F_n + \sigma_{\log F_n +}} + 10^{\log F_n - \sigma_{\log F_n -}}\right) , \tag{A.6}$$

and its uncertainty as:

$$\sigma_{F_n} = 0.5 \left(10^{\log F_n + \sigma_{\log F_n}} - 10^{\log F_n - \sigma_{\log F_n -}}\right) . \tag{A.7}$$

However, this symmetric approximation is only reasonably accurate for high signal-to-noise data.

Ideally, we should always describe the probability distribution of a transformed normal distribution by a properly transformed pdf. Unfortunately, most such distributions result in joint probability integrals that cannot be evaluated analytically – and numerical integration may be prohibitively computationally expensive when repeatedly computing likelihoods in model fitting algorithms like Galapagos. Moreover, when confronted with published measurements with asymmetric error bars, but no information as to the true underlying probability distribution, by necessity we must assume some distribution in order to proceed. To this end, we have chosen to adopt and explore the properties of a particular "asymmetric Gaussian", or skew-normal, distribution, which has proven to be amenable to analysis. For instance, if a measurement is quoted with symmetric log-normal error bars, $\log y_n \pm \sigma_{\log y,n}$, and if $\sigma_{\log y,n} \lesssim 0.1$, we can approximate the distribution in linear y-space as a skew-normal, or asymmetric Gaussian, as

described in §2.7. We define the normalized asymmetric Gaussian function as:

$$G_A(y; y_n, \sigma_{y,n\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{y,n+} + \sigma_{y,n-})} \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n+}}\right)^2\right] \text{ if } y \geq y_n \\ \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n-}}\right)^2\right] \text{ if } y < y_n \end{cases} . \quad \text{(A.8)}$$

This can also be expressed as:

$$G_A(y; y_n, \sigma_{y,n\pm}) \equiv \frac{2}{\sqrt{2\pi}(\sigma_{y,n+} + \sigma_{y,n-})} \exp\left[-\frac{1}{2}\left(\frac{y-y_n}{\sigma_{y,n\pm}}\right)^2\right] \times \Theta_{\pm}(y; y_n)\,, \quad \text{(A.9)}$$

where $\Theta_{\pm}(y; y_n)$ is the Heaviside step function:

$$\begin{aligned} \Theta_+(y; y_n) &\equiv \begin{cases} 1 \text{ if } y \geq y_n \\ 0 \text{ otherwise}\,, \end{cases} \\ \Theta_-(y; y_n) &\equiv \begin{cases} 1 \text{ if } y < y_n \\ 0 \text{ otherwise}\,. \end{cases} \end{aligned} \quad \text{(A.10)}$$

If both the probability distributions in Equation A.1 are approximated as asymmetric Gaussians (the symmetric Gaussian distribution being just a special case of G_A where $\sigma_{y+} = \sigma_{y-}$), the joint probability integral can still be expressed in terms of the convolution of two distributions, as in Equation A.5:

$$p_n = \int_{-\infty}^{\infty} \left[G_A(y; 0; \sigma_{y\mp}) * G_A(y; y_n; \sigma_{y,n\pm})\right] \delta(y - y_c(x; \vartheta_m))dy\,, \quad \text{(A.11)}$$

Note the reversal of signs in the first term of the convolution. Unfortunately, the convo-

lution of two asymmetric Gaussian functions does not, in general, yield an asymmetric Gaussian. Expanding the convolution integral:

$$\begin{aligned} &G_A(y;0;\sigma_{y\mp}) * G_A(y;y_n;\sigma_{y,n\pm}) = \\ &A\int_{-\infty}^{\infty}\left[e^{-\frac{1}{2}\left(\frac{y-y'}{\sigma_{y-}}\right)^2}\Theta_-(y';y) + e^{-\frac{1}{2}\left(\frac{y-y'}{\sigma_{y+}}\right)^2}\Theta_+(y';y)\right] \\ &\times\left[e^{-\frac{1}{2}\left(\frac{y'-y_n}{\sigma_{y,n-}}\right)^2}\Theta_-(y';y_n) + e^{-\frac{1}{2}\left(\frac{y'-y_n}{\sigma_{y,n+}}\right)^2}\Theta_+(y';y_n)\right]dy'\,, \end{aligned} \tag{A.12}$$

where:

$$A \equiv \frac{2}{\pi(\sigma_{y+}+\sigma_{y-})(\sigma_{y,n+}+\sigma_{y,n-})}\,. \tag{A.13}$$

This integral has four terms, which we denote by $I_{--}, I_{-+}, I_{+-}$and I_{++}. For example:

$$I_{--} = A\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\left(\frac{y-y'}{\sigma_{y-}}\right)^2+\left(\frac{y'-y_n}{\sigma_{y,n-}}\right)^2\right]}\Theta_-(y';y)\Theta_-(y';y_n)dy'\,. \tag{A.14}$$

Expanding the argument of the exponential, and completing the square, we find this term equals:

$$I_{--} = A\frac{\sigma_{y-}\sigma_{y,n-}}{\Sigma_{--}}e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{--}}\right)^2} \times \begin{cases} \Phi\left[\frac{\sigma_{y-}}{\sigma_{y,n-}}\frac{(y-y_n)}{\Sigma_{--}}\right] & \text{if } y < y_n, \\ \Phi\left[\frac{\sigma_{y,n-}}{\sigma_{y-}}\frac{(y_n-y)}{\Sigma_{--}}\right] & \text{if } y \geq y_n, \end{cases} \tag{A.15}$$

where:

$$\Sigma_{\pm\pm} \equiv \sqrt{\sigma_{y\pm}+\sigma_{y,n\pm}}\,, \tag{A.16}$$

is the quadrature sum of the two relevant σ's, and:

$$\Phi(z) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}x^2}dx = \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right]\,, \tag{A.17}$$

is the cumulative Gaussian distribution function. The remaining terms have a similar

form:

$$I_{-+} = A\frac{\sigma_{y-}\sigma_{y,n+}}{\Sigma_{-+}} e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{-+}}\right)^2} \times \begin{cases} 0 & \text{if } y < y_n, \\ \Phi\left[\frac{\sigma_{y-}}{\sigma_{y,n+}}\frac{(y-y_n)}{\Sigma_{-+}}\right] - \Phi\left[\frac{\sigma_{y,n+}}{\sigma_{y-}}\frac{(y_n-y)}{\Sigma_{-+}}\right] & \text{if } y \geq y_n, \end{cases} \tag{A.18}$$

$$I_{+-} = A\frac{\sigma_{y+}\sigma_{y,n-}}{\Sigma_{+-}} e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{+-}}\right)^2} \times \begin{cases} \Phi\left[\frac{\sigma_{y,n-}}{\sigma_{y+}}\frac{(y_n-y)}{\Sigma_{+-}}\right] - \Phi\left[\frac{\sigma_{y+}}{\sigma_{y,n-}}\frac{(y-y_n)}{\Sigma_{+-}}\right] & \text{if } y < y_n, \\ 0 & \text{if } y \geq y_n, \end{cases} \tag{A.19}$$

$$I_{++} = A\frac{\sigma_{y+}\sigma_{y,n+}}{\Sigma_{++}} e^{-\frac{1}{2}\left(\frac{y-y_n}{\Sigma_{++}}\right)^2} \times \begin{cases} 1 - \Phi\left[\frac{\sigma_{y,n+}}{\sigma_{y+}}\frac{(y_n-y)}{\Sigma_{++}}\right] & \text{if } y < y_n, \\ 1 - \Phi\left[\frac{\sigma_{y+}}{\sigma_{y,n+}}\frac{(y-y_n)}{\Sigma_{++}}\right] & \text{if } y \geq y_n. \end{cases} \tag{A.20}$$

So, the convolution of two asymmetric Gaussian functions is the sum of these four terms:

$$G_A(y; 0; \sigma_{y\mp}) * G_A(y; y_n; \sigma_{y,n\pm}) = I_{--} + I_{-+} + I_{+-} + I_{++}\,. \tag{A.21}$$

A.2 An Algorithm for Approximating the Convolution of Two Asymmetric Gaussians

As noted before, the result of the convolution of two asymmetric Gaussian functions is not, in general, an asymmetric Gaussian function. We have extensively explored the properties of these convolution integrals for a wide range of values for $(\sigma_{y\pm}, \sigma_{y,n\pm})$, and found that an asymmetric Gaussian function of the form $G_A(y; y_n + \delta_{y,n}, \Sigma_{y,n\pm})$ is a reasonable approximation, where:

$$\Sigma_{y,n\pm} \equiv \sqrt{\sigma_{y\mp}^2 + \sigma_{y,n\pm}}\,, \tag{A.22}$$

and where the peak of the asymmetric Gaussian is shifted by some amount $\delta_{y,n}$ from the data point's most probable value, y_n. The joint probability of the model and the

data point for a 1D statistic, in this approximation, is then:

$$p_n \approx G_A(y_c(x;\vartheta_m); y_n + \delta_{y,n}; \Sigma_{y,n\pm})\,. \tag{A.23}$$

In other words, the likelihood has the form of a χ^2-like statistic, but with the data point shifted from its nominal value by $\delta_{y,n}$, and its plus and minus 1σ error bars expanded in quadrature by the minus and plus components of the model slop, respectively. But what is $\delta_{y,n}$?

For a given set of values for $(\sigma_{y\pm}, \sigma_{y,n\pm})$, we estimated $\delta_{y,n}$ numerically, by finding the value that maximized the cross-correlation between the explicit form of the convolution integral (Equation A.21) and the asymmetric Gaussian approximation (Equation A.23). We tabulated $\delta_{y,n}$ for a wide range of values of $(\sigma_{y\pm}, \sigma_{y,n\pm})$, and searched for patterns in that four-dimensional space that might allow for an analytic approximation. We found that the magnitude of the shift scales as the maximum $\sigma_{\max} = \max\{\sigma_{y\pm}, \sigma_{y,n\pm}\}$, but also as a complicated function of various ratios of the four σ values.

We begin by identifying the maximum and minimum values of both the intrinsic and model σ's, and the maximum of all four values:

$$\begin{aligned}
\sigma_L &\equiv \max\{\sigma_{y+}, \sigma_{y-}\}\,, \\
\sigma_S &\equiv \min\{\sigma_{y+}, \sigma_{y-}\}\,, \\
\sigma_{n,L} &\equiv \max\{\sigma_{y,n+}, \sigma_{y,n-}\}\,, \\
\sigma_{n,S} &\equiv \min\{\sigma_{y,n+}, \sigma_{y,n-}\}\,, \\
\sigma_{\max} &\equiv \max\{\sigma_L, \sigma_{i,L}\}\,.
\end{aligned} \tag{A.24}$$

We define the quantity ξ as the sum of the ratios of the smaller and larger σ's:

$$\xi \equiv \frac{\sigma_S}{\sigma_L} + \frac{\sigma_{n,S}}{\sigma_{n,L}}\,, \tag{A.25}$$

and the quantity η as the difference of the ratios:

$$\eta \equiv \begin{cases} \frac{\sigma_{n,S}}{\sigma_{n,L}} - \frac{\sigma_S}{\sigma_L} & \text{if } \sigma_{n,L} < \sigma_L \\ \frac{\sigma_S}{\sigma_L} - \frac{\sigma_{n,S}}{\sigma_{n,L}} & \text{if } \sigma_{n,L} \geq \sigma_L\,. \end{cases} \tag{A.26}$$

We also define the ratio r:

$$r \equiv \frac{\min\left\{\sigma_L, \sigma_{n,L}\right\}}{\max\left\{\sigma_L, \sigma_{n,L}\right\}}\,. \tag{A.27}$$

Next, we transform the quantities ξ and η:

$$\xi' = \begin{cases} \xi & \text{if } \xi \leq 1 \\ 2 - \xi & \text{if } \xi > 1 \end{cases} \tag{A.28}$$

and:

$$\eta' = \begin{cases} 0 & \text{if } \xi' = 0 \\ 2\xi' \left[\frac{1}{2}\frac{\eta}{\xi'} + 1\right]^{n(r)} - \xi' & \text{otherwise}\,, \end{cases} \tag{A.29}$$

where the exponent $n(r)$ is the empirical function:

$$n(r) \equiv r^{-0.4087}\,. \tag{A.30}$$

Next, we compute an intermediate shift δ_*:

$$\delta_* = \sigma_{\max} N(r) \left[f(\xi) g(\eta') + h(\xi)\right]\,, \tag{A.31}$$

where N, f, g and h are empirical functions:

$$N(r) = -0.5326r^2 + 1.5307r + 0.0019 \tag{A.32}$$

$$f(\xi) = \begin{cases} 0 & \text{if } \xi = 0\,, \\ 0.2454\xi^{-1.1452} & \text{if } \xi \leq 1\,, \\ 0.2454\xi^{-0.5203} & \text{if } \xi > 1\,, \end{cases} \tag{A.33}$$

$$g(\eta') = \eta'^2 \tag{A.34}$$

$$h(\xi) = -0.042\xi^2 - 0.1602\xi + 0.4884\,. \tag{A.35}$$

We now must consider various cases. First, consider the case where one of the two distributions is symmetric, i.e., $\sigma_{y+} = \sigma_{y-}$ or $\sigma_{y,n+} = \sigma_{y,n-}$. Then, the magnitude of the shift $|\delta_{y,n}| = \delta_*$. The direction of the shift, which we denote by $i = \{-1, 1\}$, is determined by the direction of the largest σ of the asymmetric Gaussian:

$$i = \begin{cases} +1 & \text{if } \sigma_{n,L} = \sigma_{y,n+} \text{ or } \sigma_L = \sigma_{y-} \\ -1 & \text{if } \sigma_{\max} = \sigma_{y,n-} \text{ or } \sigma_{\max} = \sigma_{y+}\,, \end{cases} \tag{A.36}$$

and $\delta_{y,n} = i \times \delta_*$.

If both distributions are asymmetric, there are two cases. The first case applies when the larger σ's of the slop and the intrinsic pdf are in opposite directions, i.e., when $\sigma_L = \sigma_{y-}$ and $\sigma_{n,L} = \sigma_{y,n+}$, or $\sigma_L = \sigma_{y+}$ and $\sigma_{n,L} = \sigma_{y,n-}$. In this case, as in the symmetric case above, the magnitude of the shift $|\delta_{y,n}| = \delta_*$, and the direction of the shift is:

$$i = \begin{cases} +1 & \text{if } \sigma_{\max} = \sigma_{y,n+} \text{ or } \sigma_{\max} = \sigma_{y-} \\ -1 & \text{if } \sigma_{\max} = \sigma_{y,n-} \text{ or } \sigma_{\max} = \sigma_{y+}\,, \end{cases} \tag{A.37}$$

and $\delta_{y,n} = i \times \delta_*$.

The second case applies when the larger σ's of the slop and the intrinsic pdf are in

the same direction, i.e., when $\sigma_L = \sigma_{y+}$ and $\sigma_{n,L} = \sigma_{y,n+}$, or $\sigma_L = \sigma_{y-}$ and $\sigma_{n,L} = \sigma_{y,n-}$. Here, the shift $\delta_{y,n}$ is obtained by multiplying $i \times \delta_*$ by an additional empirical function:

$$\delta_{y,n} = i \times \delta_* \times \sin\left(\frac{\pi}{2}\frac{\eta'}{\xi'}\right) \times \begin{cases} \xi^{0.7413} & \text{if } \xi \leq 1 \\ \xi^{-0.1268} & \text{if } \xi > 1\,, \end{cases} \tag{A.38}$$

where i is given, as before, by Equation A.37. Note that, since the sinusoidal term in Equation A.38 may take negative values, the actual direction of the shift may be opposite in sign to i.[1]

[1]There is one specific case which requires special treatment. If both distributions are asymmetric, but if, by chance, $\sigma_{y,n-} = \sigma_{y-}$ exactly, and $\sigma_{y+} < \sigma_{y,n+} < \sigma_{y,n-}$, then the sign of i from Equation A.37 must be reversed.

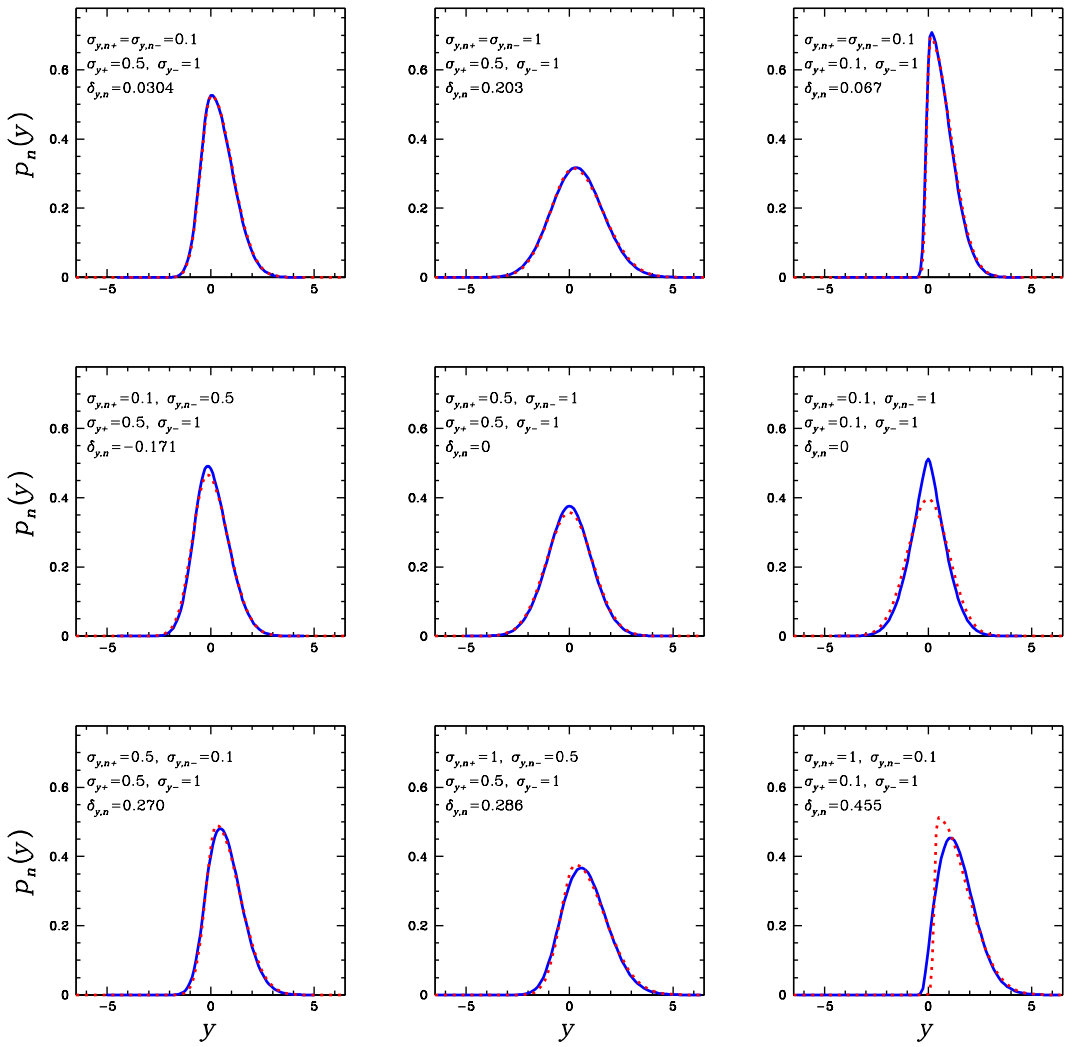

Figure A.1 Asymmetric Gaussian approximations of the convolution of two asymmetric Gaussians, $G_A(y;0,\sigma_{y\mp}) * G_A(y;0,\sigma_{y,n\pm})$. Solid blue curves are the actual convolved distributions. Dotted red curves are the asymmetric Gaussian approximations $G_A(y;\delta_{y,n},\Sigma_{y,n\pm})$, where $\Sigma^2_{y,n\pm} \equiv \sigma^2_{y\mp} + \sigma^2_{y,n\pm}$.

Appendix B

Evaluating the Joint Probability of Two Asymmetric Gaussian Distributions in Two Dimensions

We proceed, as in §2.2.1, by approximating the model curve $y_c(x;\vartheta_m)$ by a line tangent to the asymmetric convolved (and shifted) error "ellipse", taking care to keep track of which quadrant the tangent point lies in. However, since the tangent line will, in general, cross through three quadrants of the asymmetric convolved ellipse, the joint probability integral over the range $[-\infty,\infty]$ of Equation 2.14 must be broken into three segments.

Consider, for example, the case shown in Figure B.1: a model distribution about the curve $y_c(x;\vartheta_m)$ with asymmetric Gaussian slops $(\sigma_{x\pm},\sigma_{y\pm})$, and a data point at (x_n,y_n) with asymmetric Gaussian intrinsic uncertainties $(\sigma_{x,n\pm},\sigma_{y,n\pm})$. As discussed in §2.7.2 and Appendix A, evaluation of the joint probability integral (Equation 2.6) requires a convolution of these two 2D asymmetric Gaussian distributions. The resulting convolved distribution can be approximated as a 2D asymmetric Gaussian, with 1σ widths $(\Sigma_{x,n\pm},\Sigma_{y,n\pm})$, and a centroid at $(x_n+\delta_{x,n},y_n+\delta_{y,n})$, where $\Sigma^2_{x,n\pm}=\sigma^2_{x\mp}+\sigma^2_{x,n\pm}$, $\Sigma^2_{y,n\pm}=\sigma^2_{y\mp}+\sigma^2_{y,n\pm}$, and where the centroid shifts $(\delta_{x,n},\delta_{y,n})$ are given by the algorithm in Appendix A. In the case shown in Figure B.1, the tangent point $(x_{t,n},y_{t,n})$ lies in quadrant 1 of the asymmetric convolved error ellipse, where the curve has a slope $m_{t,n}$, and so, for the TRF statistic, the joint probability will be proportional to the factor

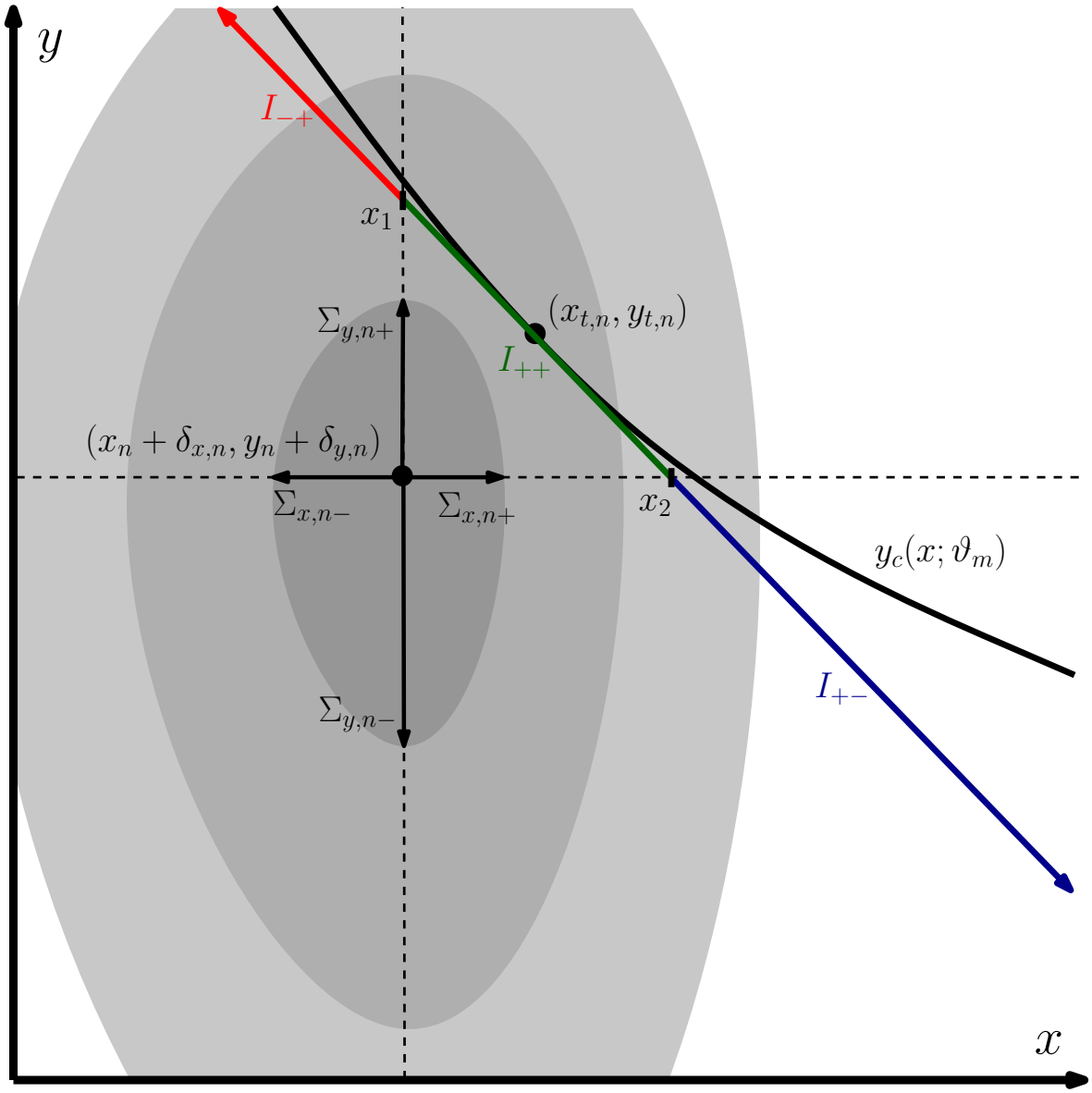

Figure B.1 Geometry of the 2D joint probability integral for asymmetric probability distributions. The convolution of an asymmetric model distribution and the asymmetric intrinsic uncertainty of a data point is approximated by a 2D asymmetric Gaussian, with 1σ widths $(\Sigma_{x,n\pm}, \Sigma_{y,n\pm})$, and with a centroid at $(x_n + \delta_{x,n}, y_n + \delta_{y,n})$. The model curve $y_c(x; \vartheta_m)$ is approximated by a line tangent to the convolved ellipse at point $(x_{t,n}, y_{t,n})$. The joint probability integral must be broken into three segments. The integral I_{-+} (red) through quadrant 2 has limits $[-\infty, x_1]$. The integral I_{++} (green) through quadrant 1 has limits $[x_1, x_2]$. The integral I_{+-} (blue) through quadrant 4 has limits $[x_2, \infty]$.

(see §2.3):

$$\left(\frac{du_n}{dx}\right)_{++} = \frac{m_{t,n}^2\Sigma_{x,n+}^2 + \Sigma_{y,n+}^2}{\sqrt{m_{t,n}^2\Sigma_{x,n+}^4 + \Sigma_{y,n+}^4}}\,. \tag{B.1}$$

In general, the factor $\left(\frac{du_n}{dx}\right)_{\pm\mp}$ is determined by which quadrant the tangent point lies in:

$$\left(\frac{du_n}{dx}\right)_{\pm\mp} = \frac{m_{t,n}^2\Sigma_{x,n\pm}^2 + \Sigma_{y,n\mp}^2}{\sqrt{m_{t,n}^2\Sigma_{x,n\pm}^4 + \Sigma_{y,n\mp}^4}}\,, \tag{B.2}$$

where $\left(\frac{du_n}{dx}\right)_{++}$ corresponds to quadrant 1, $\left(\frac{du_n}{dx}\right)_{-+}$ to quadrant 2, $\left(\frac{du_n}{dx}\right)_{--}$ to quadrant 3, and $\left(\frac{du_n}{dx}\right)_{+-}$ to quadrant 4, and where the first subscript indicates the choice of $\Sigma_{x,n\pm}$, and the second indicates the choice of $\Sigma_{y,n\mp}$ in Equation B.2.

The joint probability integral of Equation 2.14 must be broken into three segments, with limits $[-\infty, x_1]$, $[x_1, x_2]$ and $[x_2, \infty]$, where $x_1 = x_n + \delta_{x,n}$, and where x_2 is the point where the tangent line $y_{t,n} + m_{t,n}(x - x_{t,n}) = y_n + \delta_{y,n}$. The first segment of the integral, through quadrant 2, is:

$$\begin{aligned} I_{-+} &= \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n-}\Sigma_{y,n+}}{\pi(\Sigma_{x,n+} + \Sigma_{x,n-})(\Sigma_{y,n+} + \Sigma_{y,n-})} \frac{\Phi(z_{-+}(x_1))}{\sqrt{m_{t,n}^2\Sigma_{x,n-}^2 + \Sigma_{y,n+}^2}} \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n} - y_n - \delta_{y,n} - m_{t,n}(x_{t,n} - x_n - \delta_{x,n})}{\sqrt{m_{t,n}^2\Sigma_{x,n-}^2 + \Sigma_{y,n+}^2}}\right]^2\right\}, \end{aligned} \tag{B.3}$$

where $\Phi(z)$ is the cumulative Gaussian function:

$$\Phi(z) \equiv \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx\,, \tag{B.4}$$

and where the limit of integration x_1 is transformed to:

$$z_{-+}(x_1) = \frac{\Sigma_{y,n+}^2(x_1 - x_n - \delta_{x,n}) + m_{t,n}^2\Sigma_{x,n-}^2\left[x_1 - x_{t,n} - (y_n + \delta_{y,n} - y_{t,n})/m_{t,n}\right]}{\Sigma_{x,n-}\Sigma_{y,n+}\sqrt{m_{t,n}^2\Sigma_{x,n-}^2 + \Sigma_{y,n+}^2}}\,. \tag{B.5}$$

This transformation of the limit from x to $z_{\pm\mp}(x)$ is generally applicable to integrations through all quadrants. In general:

$$z_{\pm\mp}(x) = \frac{\Sigma^2_{y,n\mp}(x - x_n - \delta_{x,n}) + m^2_{t,n}\Sigma^2_{x,n\pm}\left[x - x_{t,n} - (y_n + \delta_{y,n} - y_{t,n})/m_{t,n}\right]}{\Sigma_{x,n\pm}\Sigma_{y,n\mp}\sqrt{m^2_{t,n}\Sigma^2_{x,n\pm} + \Sigma^2_{y,n\mp}}}, \quad \text{(B.6)}$$

where $z_{++}(x)$ corresponds to integration through quadrant 1, z_{-+} to quadrant 2, z_{--} to quadrant 3, and z_{+-} to quadrant 4, and where the first subscript of $z_{\pm\mp}(x)$ indicates the choice of $\Sigma_{x,n\pm}$, and the second indicates the choice of $\Sigma_{y,n\mp}$ in Equation B.6.

The second segment of the integral, through quadrant 1 between the limits $[x_1, x_2]$, is:

$$\begin{aligned} I_{++} &= \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n+}\Sigma_{y,n+}}{\pi(\Sigma_{x,n+} + \Sigma_{x,n-})(\Sigma_{y,n+} + \Sigma_{y,n-})} \frac{[\Phi(z_{++}(x_2)) - \Phi(z_{++}(x_1))]}{\sqrt{m^2_{t,n}\Sigma^2_{x,n+} + \Sigma^2_{y,n+}}} \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n} - y_n - \delta_{y,n} - m_{t,n}(x_{t,n} - x_n - \delta_{x,n})}{\sqrt{m^2_{t,n}\Sigma^2_{x,n+} + \Sigma^2_{y,n+}}}\right]^2\right\} \\ &\quad . \end{aligned} \quad \text{(B.7)}$$

Finally, the third segment of the integral, through quadrant 4, is:

$$\begin{aligned} I_{+-} &= \left(\frac{du_n}{dx}\right)_{++} \frac{2\Sigma_{x,n+}\Sigma_{y,n-}}{\pi(\Sigma_{x,n+} + \Sigma_{x,n-})(\Sigma_{y,n+} + \Sigma_{y,n-})} \frac{[1 - \Phi(z_{+-}(x_2))]}{\sqrt{m^2_{t,n}\Sigma^2_{x,n+} + \Sigma^2_{y,n-}}} \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y_{t,n} - y_n - \delta_{y,n} - m_{t,n}(x_{t,n} - x_n - \delta_{x,n})}{\sqrt{m^2_{t,n}\Sigma^2_{x,n+} + \Sigma^2_{y,n-}}}\right]^2\right\}. \end{aligned} \quad \text{(B.8)}$$

The total joint probability integral is then the sum of these three terms, $p_n = I_{-+} + I_{++} + I_{+-}$.

This procedure is easily generalized to cases where the tangent point falls in any of the other quadrants. In general, the integral will always consist of three segments

through three different quadrants, one with limits $[-\infty, x_1]$, one with limits $[x_1, x_2]$, and one with limits $[x_2, \infty]$. The factor $\left(\frac{du_n}{dx}\right)_{\pm\mp}$ (Equation B.2) is computed according to the quadrant the tangent point lies in, and is the same for all three integrals. Similarly, the normalizing factor $2/(\pi(\Sigma_{x,n+} + \Sigma_{x,n-})(\Sigma_{y,n+} + \Sigma_{y,n-}))$ will be the same for all three integrals. The first integral will take the form of Equation B.3, but with a different choice of $(\Sigma_{x,n\pm}, \Sigma_{y,n\mp})$ in that expression and in the expression for the transformed limit $z_{\pm\mp}(x_1)$ (Equation B.6). Likewise, the second integral will take the form of Equation B.7, and the third the form of Equation B.8, each expressed in terms of the values of $(\Sigma_{x,n\pm}, \Sigma_{y,n\mp})$ appropriate to the quadrant that that segment of the integral lies in. For example, if the tangent point lies in quadrant 4, all three integrals will be proportional to the factor $\left(\frac{du_n}{dx}\right)_{+-}$. The first integral, from $[-\infty, x_1]$, will be through quadrant 3, and have the form of Equation B.3, but with $(\Sigma_{x,n-}, \Sigma_{y,n-})$. The second integral, from $[x_1, x_2]$, will be through quadrant 4, and have the form of Equation B.7, but with $(\Sigma_{x,n+}, \Sigma_{y,n-})$. And the third integral, from $[x_2, \infty]$, will through quadrant 1, and have the form of Equation B.8, but with $(\Sigma_{x,n+}, \Sigma_{y,n+})$.

Appendix C

Computing 2D Model Distribution Envelopes for Plotting

In general, for a best-fit curve $y_c(x;\vartheta_m)$, described by M parameters $\{\vartheta_m\}$, and with best-fit slop parameters (σ_x, σ_y), the $n\sigma$ slop envelope is defined as the locus of points that is tangent to the set of ellipses of widths $(n\sigma_x, n\sigma_y)$, centered on each point along the curve. Consider a point on the curve, $(x, y_c(x;\vartheta_m))$, at which the curve has a slope $m_c(x;\vartheta_m) = \frac{d}{dx}y_c(x;\vartheta_m)$. An $n\sigma$ ellipse centered on that point is described by the equation:

$$\frac{\Delta x_{n\sigma}^2}{\sigma_x^2} + \frac{\Delta y_{n\sigma}^2}{\sigma_y^2} = n^2\,, \tag{C.1}$$

where $(\Delta x_{n\sigma}, \Delta y_{n\sigma})$ are the distances from the point $(x, y_c(x;\vartheta_m))$. Differentiating Equation C.1 with respect to x, and solving for $\frac{dy}{dx}$ gives the slope of the ellipse at $(\Delta x_{n\sigma}, \Delta y_{n\sigma})$:

$$m = -\frac{\sigma_y^2}{\sigma_x^2}\frac{\Delta x_{n\sigma}}{\Delta y_{n\sigma}}\,. \tag{C.2}$$

We are interested in the point on the ellipse $(\Delta x_{n\sigma_c}, \Delta y_{n\sigma_c})$ at which the slope is equal to that of the curve $m_c(x;\vartheta_m)$. From Equation C.2, this is equivalent to the point for which:

$$\Delta x_{n\sigma_c} = -m_c(x,\vartheta_m)\frac{\sigma_x^2}{\sigma_y^2}\Delta y_{n\sigma_c}\,. \tag{C.3}$$

Substituting Equation C.3 into Equation C.1, and solving for $\Delta y_{n\sigma_c}$ yields:

$$\Delta y_{n\sigma_c} = \pm\frac{n\sigma_y}{\sqrt{1+\frac{m_c^2(x;\vartheta_m)\sigma_x^2}{\sigma_y^2}}}\,. \tag{C.4}$$

Together, Equations C.4 & C.3 describe the offset of the $n\sigma$ slop envelope from each point on the model curve. The model distribution envelopes shown in Figures 3.8, 3.10 & 3.12 were computed using this technique.

If the fitted uncertainties in the model curve parameters themselves are sufficiently great, one may wish to plot combined slop and model parameter uncertainty envelopes. There is no ideal technique for plotting such envelopes; one is simply to approximate them by projecting the uncertainty of each parameter to first order into the y-direction, and adding those projected uncertainties in quadrature to the slop envelopes given by Equations C.3 & C.4:

$$\begin{aligned} \Delta y_{n\sigma_c;\, n\sigma_{\vartheta_m}} &= \left\{ \Delta y_{n\sigma_c}^2 + \sum_{k=1}^{M} \left[n\sigma_{\vartheta_k} \times \frac{\partial}{\partial \vartheta_k} y_c(x;\vartheta_m) \right]^2 \right\}^{1/2} , \\ \Delta x_{n\sigma_c;\, n\sigma_{\vartheta_m}} &= \Delta x_{n\sigma_c} . \end{aligned} \tag{C.5}$$

Note that Equations C.3, C.4 & C.5 assume that both the slop and the uncertainties in the model parameters are described by symmetric Gaussian functions. For asymmetric and/or non-Gaussian distributions, some care must be taken to ensure that each offset is of the proper magnitude and in the proper direction. Finally, we note that it is inappropriate to plot combined slop and model parameter uncertainty envelopes together with data. *Only best-fit slop envelopes should be plotted with data.* However, the combined envelopes can be plotted separately, and are useful for indicating the total uncertainty as a function of x. However, in the case of the fits presented in Chapter 3, we do not do so, since the two envelopes are effectively indistinguishable.

Bibliography

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Chevalier, R. A. & Li, Z. 1999, ApJ, 520, L29

Chornock, R., Li, W., & Filippenko, A. V. 2009a, GRB Coordinates Network, 8979, 1

Chornock, R., Perley, D. A., Cenko, S. B., Bloom, J. S., Cobb, B., & Prochaska, J. X. 2009b, GRB Coordinates Network, 8994, 1

Cobb, B. E. 2009, GRB Coordinates Network, 9008, 1

D'Agostini, G. 2005, ArXiv Physics e-prints

de Ugarte Postigo, A., Gorosabel, J., Sota, A., Castro-Tirado, A. J., McBreen, S., & Zapatero-Osorio, M. R. 2009, GRB Coordinates Network, 8999, 1

Draine, B. T. 2000, ApJ, 532, 273

Draine, B. T. & Lee, H. M. 1984, ApJ, 285, 89

Evans, P. A., Beardmore, A. P., Page, K. L., Osborne, J. P., O'Brien, P. T., Willingale, R., Starling, R. L. C., Burrows, D. N., Godet, O., Vetere, L., Racusin, J., Goad, M. R., Wiersema, K., Angelini, L., Capalbi, M., Chincarini, G., Gehrels, N., Kennea, J. A., Margutti, R., Morris, D. C., Mountford, C. J., Pagani, C., Perri, M., Romano, P., & Tanvir, N. 2009, MNRAS, 397, 1177

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., & Fukugita, M. 2006, AJ, 132, 117

Fitzpatrick, E. L. & Massa, D. 1988, ApJ, 328, 734

Freedman, W. L., Kennicutt, R. C., & Mould, J. R. 2010, Highlights of Astronomy, 15, 1

Fruchter, A., Krolik, J. H., & Rhoads, J. E. 2001, ApJ, 563, 597

Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279

Granot, J., Piran, T., & Sari, R. 2000, ApJ, 534, L163

Granot, J. & Sari, R. 2002, ApJ, 568, 820

Greenberg, J. M. & Chlewicki, G. 1983, ApJ, 272, 563

Hecht, J. H. 1986, ApJ, 305, 817

Jenniskens, P. & Greenberg, J. M. 1993, A&A, 274, 439

Katz, J. I. 1994, ApJ, 432, L107

Lee, B. C., Tucker, D. L., Vanden Berk, D. E., Yanny, B., Reichart, D. E., Adelman, J., Chen, B., Harvanek, M., Henden, A., Ivezić, Ž., Kleinman, S., Lamb, D., Long, D., McMillan, R., Newman, P. R., Nitta, A., Palunas, P., Schneider, D. P., Snedden, S., York, D., Briggs, J. W., Brinkmann, J., Csabai, I., Hennessy, G. S., Kent, S., Lupton, R., Newberg, H. J., & Stoughton, C. 2001, ApJ, 561, 183

Mao, J., Beardmore, A. P., Evans, P. A., Guidorzi, C., Hoversten, E. A., Krimm, H. A., Margutti, R., O'Brien, P. T., Page, K. L., Romano, P., Sakamoto, T., Tagliaferri, G., & Ziaeepour, H. 2009, GRB Coordinates Network, 8980, 1

Mao, J. & Margutti, R. 2009, GRB Coordinates Network, 8991, 1

Mészáros, P. 2002, ARA&A, 40, 137

Mészáros, P. & Rees, M. J. 1997, ApJ, 476, 232

Morgan, A. N., Bloom, J. S., Perley, D. A., & Starr, D. 2009, GRB Coordinates Network, 8995, 1

Morrison, R. & McCammon, D. 1983, ApJ, 270, 119

Pearson, K. 1896, Philos. Trans. Royal Soc. London Ser. A, 187, 253

Perley, D. A. 2009, GRB Coordinates Network, 8997, 1

Perley, D. A., Metzger, B. D., Granot, J., Butler, N. R., Sakamoto, T., Ramirez-Ruiz, E., Levan, A. J., Bloom, J. S., Miller, A. A., Bunker, A., Chen, H.-W., Filippenko, A. V., Gehrels, N., Glazebrook, K., Hall, P. B., Hurley, K. C., Kocevski, D., Li, W., Lopez, S., Norris, J., Piro, A. L., Poznanski, D., Prochaska, J. X., Quataert, E., & Tanvir, N. 2009, ApJ, 696, 1871

Piran, T. 1999, Phys. Rep., 314, 575

Reichart, D., Nysewander, M., Moran, J., Bartelme, J., Bayliss, M., Foster, A., Clemens, J. C., Price, P., Evans, C., Salmonson, J., Trammell, S., Carney, B., Keohane, J., & Gotwals, R. 2005, Nuovo Cimento C Geophysics Space Physics C, 28, 767

Reichart, D. E. 2001, ApJ, 553, 235

Reichart, D. E. & Price, P. A. 2002, ApJ, 565, 174

Rodgers, J. L. & Nicewander, W. A. 1988, American Statistician, 42, 59

Rybicki, G. B. & Lightman, A. P. 1979, Radiative Processes in Astrophysics (New York, NY: John Wiley & Sons)

Sari, R., Piran, T., & Halpern, J. P. 1999, ApJ, 519, L17

Sari, R., Piran, T., & Narayan, R. 1998, ApJ, 497, L17+

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Songaila, A. 2004, AJ, 127, 2598

Starling, R. L. C., van der Horst, A. J., Rol, E., Wijers, R. A. M. J., Kouveliotou, C., Wiersema, K., Curran, P. A., & Weltevrede, P. 2008, ApJ, 672, 433

Totani, T., Kawai, N., Kosugi, G., Aoki, K., Yamada, T., Iye, M., Ohta, K., & Hattori, T. 2006, PASJ, 58, 485

Updike, A. C., Klose, S., Clemens, C., & Greiner, J. 2009, GRB Coordinates Network, 8983, 1

Valencic, L. A., Clayton, G. C., & Gordon, K. D. 2004, ApJ, 616, 912

van de Hulst, H. C. 1957, Light Scattering by Small Particles (New York, NY: John Wiley & Sons)

Woosley, S. E. & Bloom, J. S. 2006, ARA&A, 44, 507